

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/12, C07K 14/47, C12N 15/10, 15/11</b>		<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 99/06553</b> <b>(43) International Publication Date:</b> 11 February 1999 (11.02.99)
<b>(21) International Application Number:</b> PCT/IB98/01237 <b>(22) International Filing Date:</b> 31 July 1998 (31.07.98) <b>(30) Priority Data:</b> 08/905,051 1 August 1997 (01.08.97) US <b>(71) Applicant (for all designated States except US):</b> GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> DUMAS MILNE EDWARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire de Tours, F-75006 Paris (FR). DUCLERT, Aymeric [FR/FR]; 6 ter, rue Victorine, F-94100 Saint-Maur (FR). LACROIX, Bruno [FR/FR]; 9, route de Vourles, F-69230 Saint-Genis Laval (FR). <b>(74) Agents:</b> MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).			<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> 5' ESTs FOR SECRETED PROTEINS EXPRESSED IN VARIOUS TISSUES			
<b>(57) Abstract</b> <p>The sequences of 5' ESTs derived from mRNAs encoding secreted proteins are disclosed. The 5' ESTs may be to obtain cDNAs and genomic DNAs corresponding to the 5' ESTs. The 5' ESTs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. Upstream regulatory sequences may also be obtained using the 5' ESTs. The 5' ESTs may also be used to design expression vectors and secretion vectors.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## 5' ESTs FOR SECRETED PROTEINS EXPRESSED IN VARIOUS TISSUES

### Background of the Invention

The estimated 50,000-100,000 genes scattered along the human chromosomes offer tremendous promise for the understanding, diagnosis, and treatment of human diseases. In addition, probes capable of specifically hybridizing to loci distributed throughout the human genome find applications in the construction of high resolution chromosome maps and in the identification of individuals.

In the past, the characterization of even a single human gene was a painstaking process, requiring years of effort. Recent developments in the areas of cloning vectors, DNA sequencing, and computer technology have merged to greatly accelerate the rate at which human genes can be isolated, sequenced, mapped, and characterized. Cloning vectors such as yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) are able to accept DNA inserts ranging from 300 to 1000 kilobases (kb) or 100-400 kb in length respectively, thereby facilitating the manipulation and ordering of DNA sequences distributed over great distances on the human chromosomes. Automated DNA sequencing machines permit the rapid sequencing of human genes. Bioinformatics software enables the comparison of nucleic acid and protein sequences, thereby assisting in the characterization of human gene products.

Currently, two different approaches are being pursued for identifying and characterizing the genes distributed along the human genome. In one approach, large fragments of genomic DNA are isolated, cloned, and sequenced. Potential open reading frames in these genomic sequences are identified using bioinformatics software. However, this approach entails sequencing large stretches of human DNA which do not encode proteins in order to find the protein encoding sequences scattered throughout the genome. In addition to requiring extensive sequencing, the bioinformatics software may mischaracterize the genomic sequences obtained. Thus, the software may produce false positives in which non-coding DNA is mischaracterized as coding DNA or false negatives in which coding DNA is mislabeled as non-coding DNA.

An alternative approach takes a more direct route to identifying and characterizing human genes. In this approach, complementary DNAs (cDNAs) are synthesized from isolated messenger RNAs (mRNAs) which encode human proteins. Using this approach,

sequencing is only performed on DNA which is derived from protein coding portions of the genome. Often, only short stretches of the cDNAs are sequenced to obtain sequences called expressed sequence tags (ESTs). The ESTs may then be used to isolate or purify extended cDNAs which include sequences adjacent to the EST sequences. The extended cDNAs may  
5 contain all of the sequence of the EST which was used to obtain them or only a portion of the sequence of the EST which was used to obtain them. In addition, the extended cDNAs may contain the full coding sequence of the gene from which the EST was derived or, alternatively, the extended cDNAs may include portions of the coding sequence of the gene from which the EST was derived. It will be appreciated that there may be several extended  
10 cDNAs which include the EST sequence as a result of alternate splicing or the activity of alternative promoters.

In the past, these short EST sequences were often obtained from oligo-dT primed cDNA libraries. Accordingly, they mainly corresponded to the 3' untranslated region of the mRNA. In part, the prevalence of EST sequences derived from the 3' end of the mRNA is a  
15 result of the fact that typical techniques for obtaining cDNAs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs. (Adams *et al.*, *Nature* 377:3-174, 1996; Hillier *et al.*, *Genome Res.* 6:807-828, 1996).

In addition, in those reported instances where longer cDNA sequences have been obtained, the reported sequences typically correspond to coding sequences and do not include  
20 the full 5' untranslated region of the mRNA from which the cDNA is derived. Such incomplete sequences may not include the first exon of the mRNA, particularly in situations where the first exon is short. Furthermore, they may not include some exons, often short ones, which are located upstream of splicing sites. Thus, there is a need to obtain sequences derived from the 5' ends of mRNAs.

25 While many sequences derived from human chromosomes have practical applications, approaches based on the identification and characterization of those chromosomal sequences which encode a protein product are particularly relevant to diagnostic and therapeutic uses. Of the 50,000-100,000 protein coding genes, those genes encoding proteins which are secreted from the cell in which they are synthesized, as well as the secreted proteins  
30 themselves, are particularly valuable as potential therapeutic agents. Such proteins are often



involved in cell to cell communication and may be responsible for producing a clinically relevant response in their target cells.

In fact, several secretory proteins, including tissue plasminogen activator, G-CSF, GM-CSF, erythropoietin, human growth hormone, insulin, interferon- $\alpha$ , interferon- $\beta$ , interferon- $\gamma$ , and interleukin-2, are currently in clinical use. These proteins are used to treat a wide range of conditions, including acute myocardial infarction, acute ischemic stroke, anemia, diabetes, growth hormone deficiency, hepatitis, kidney carcinoma, chemotherapy induced neutropenia and multiple sclerosis. For these reasons, extended cDNAs encoding secreted proteins or portions thereof represent a particularly valuable source of therapeutic agents. Thus, there is a need for the identification and characterization of secreted proteins and the nucleic acids encoding them.

In addition to being therapeutically useful themselves, secretory proteins include short peptides, called signal peptides, at their amino termini which direct their secretion. These signal peptides are encoded by the signal sequences located at the 5' ends of the coding sequences of genes encoding secreted proteins. Because these signal peptides will direct the extracellular secretion of any protein to which they are operably linked, the signal sequences may be exploited to direct the efficient secretion of any protein by operably linking the signal sequences to a gene encoding the protein for which secretion is desired. In addition, portions of signal sequences may also be used to direct the intracellular import of a peptide or protein of interest. This may prove beneficial in gene therapy strategies in which it is desired to deliver a particular gene product to cells other than the cell in which it is produced. Signal sequences encoding signal peptides also find application in simplifying protein purification techniques. In such applications, the extracellular secretion of the desired protein greatly facilitates purification by reducing the number of undesired proteins from which the desired protein must be selected. Thus, there exists a need to identify and characterize the 5' portions of the genes for secretory proteins which encode signal peptides.

Public information on the number of human genes for which the promoters and upstream regulatory regions have been identified and characterized is quite limited. In part, this may be due to the difficulty of isolating such regulatory sequences. Upstream regulatory sequences such as transcription factor binding sites are typically too short to be utilized as probes for isolating promoters from human genomic libraries. Recently, some approaches

have been developed to isolate human promoters. One of them consists of making a CpG island library (Cross, *et al.*, *Nature Genetics* 6: 236-244, 1994). The second consists of isolating human genomic DNA sequences containing SpeI binding sites by the use of SpeI binding protein. (Mortlock *et al.*, *Genome Res.* 6:327-335, 1996). Both of these approaches  
5 have their limits due to a lack of specificity or of comprehensiveness.

The present 5' ESTs may be used to efficiently identify and isolate upstream regulatory regions which control the location, developmental stage, rate, and quantity of protein synthesis, as well as the stability of the mRNA. (Theil, *BioFactors* 4:87-93, 1993). Once identified and characterized, these regulatory regions may be utilized in gene therapy or  
10 protein purification schemes to obtain the desired amount and locations of protein synthesis or to inhibit, reduce, or prevent the synthesis of undesirable gene products.

In addition, ESTs containing the 5' ends of secretory protein genes may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5' coding  
15 sequences of genes encoding secretory proteins.

#### Summary of the Invention

The present invention relates to purified, isolated, or recombinant ESTs which include sequences derived from the authentic 5' ends of their corresponding mRNAs. The term  
20 "corresponding mRNA" refers to the mRNA which was the template for the cDNA synthesis which produced the 5' EST. These sequences will be referred to hereinafter as "5' ESTs." As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual 5' EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these  
25 clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus,  
30 creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately  $10^4$ - $10^6$  fold purification of the native message.

Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

As used herein, the term "recombinant" means that the 5' EST is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the 5' ESTs will represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the present invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Preferably, the enriched 5' ESTs represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More preferably, the enriched 5' ESTs represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a highly preferred embodiment, the enriched 5' ESTs represent 90% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules.

"Stringent", moderate, and "low" hybridization conditions are as defined in Example 29.

Unless otherwise indicated, a "complementary" sequence is fully complementary.

Thus, 5' ESTs in cDNA libraries in which one or more 5' ESTs make up 5% or more of the number of nucleic acid inserts in the backbone molecules are "enriched recombinant 5' ESTs" as defined herein. Likewise, 5' ESTs in a population of plasmids in which one or more 5' EST of the present invention have been inserted such that they represent 5% or more of the number of inserts in the plasmid backbone are "enriched recombinant 5' ESTs" as defined herein. However, 5' ESTs in cDNA libraries in which 5' ESTs constitute less than 5% of the number of nucleic acid inserts in the population of backbone molecules, such as libraries in

which backbone molecules having a 5' EST insert are extremely rare, are not "enriched recombinant 5' ESTs."

In particular, the present invention relates to 5' ESTs which are derived from genes encoding secreted proteins. As used herein, a "secreted" protein is one which, when expressed in a suitable host cell, is transported across or through a membrane, including transport as a result of signal peptides in its amino acid sequence. "Secreted" proteins include without limitation proteins secreted wholly (e.g. soluble proteins), or partially (e.g. receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins which are transported across the membrane of the endoplasmic reticulum.

Such 5' ESTs include nucleic acid sequences, called signal sequences, which encode signal peptides which direct the extracellular secretion of the proteins encoded by the genes from which the 5' ESTs are derived. Generally, the signal peptides are located at the amino termini of secreted proteins.

Secreted proteins are translated by ribosomes associated with the "rough" endoplasmic reticulum. Generally, secreted proteins are co-translationally transferred to the membrane of the endoplasmic reticulum. Association of the ribosome with the endoplasmic reticulum during translation of secreted proteins is mediated by the signal peptide. The signal peptide is typically cleaved following its co-translational entry into the endoplasmic reticulum.

After delivery to the endoplasmic reticulum, secreted proteins may proceed through the Golgi apparatus. In the Golgi apparatus, the proteins may undergo post-translational modification before entering secretory vesicles which transport them across the cell membrane.

The 5' ESTs of the present invention have several important applications. For example, they may be used to obtain and express cDNA clones which include the full protein coding sequences of the corresponding gene products, including the authentic translation start sites derived from the 5' ends of the coding sequences of the mRNAs from which the 5' ESTs are derived. These cDNAs will be referred to hereinafter as "full length cDNAs." These cDNAs may also include DNA derived from mRNA sequences upstream of the translation start site. The full length cDNA sequences may be used to express the proteins corresponding to the 5' ESTs. As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the cDNAs may be useful in treating or

controlling a variety of human conditions. The 5' ESTs may also be used to obtain the corresponding genomic DNA. The term "corresponding genomic DNA" refers to the genomic DNA which encodes the mRNA from which the 5' EST was derived.

Alternatively, the 5' ESTs may be used to obtain and express extended cDNAs  
5 encoding portions of the secreted protein. The portions may comprise the signal peptides of the secreted proteins or the mature proteins generated when the signal peptide is cleaved off. The portions may also comprise polypeptides having at least 10 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. Alternatively, the portions may comprise at least 15 consecutive amino acids encoded by the extended cDNAs or full length  
10 cDNAs. In some embodiments, the portions may comprise at least 25 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. In other embodiments, the portions may comprise at least 40 amino acids encoded by the extended cDNAs or full length cDNAs.

Antibodies which specifically recognize the entire secreted proteins encoded by the  
15 extended cDNAs, full length cDNAs, or fragments thereof having at least 10 consecutive amino acids, at least 15 consecutive amino acids, at least 25 consecutive amino acids, or at least 40 consecutive amino acids may also be obtained as described below. Antibodies which specifically recognize the mature protein generated when the signal peptide is cleaved may also be obtained as described below. Similarly, antibodies which specifically recognize the  
20 signal peptides encoded by the extended cDNAs or full length cDNAs may also be obtained.

In some embodiments, the extended cDNAs obtained using the 5' ESTs include the signal sequence. In other embodiments, the extended cDNAs obtained using the 5' ESTs may include the full coding sequence for the mature protein (*i.e.* the protein generated when the signal polypeptide is cleaved off). In addition, the extended cDNAs obtained using the 5'  
25 ESTs may include regulatory regions upstream of the translation start site or downstream of the stop codon which control the amount, location, or developmental stage of gene expression.

As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the extended cDNAs or full length cDNAs obtained using the 5'  
30 ESTs may be useful in treating or controlling a variety of human conditions.

The 5' ESTs (or cDNAs or genomic DNAs obtained therefrom) may be used in forensic procedures to identify individuals or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal expression of the genes corresponding to the 5' ESTs. In addition, the present invention is useful for constructing a high resolution map of the human chromosomes.

The present invention also relates to secretion vectors capable of directing the secretion of a protein of interest. Such vectors may be used in gene therapy strategies in which it is desired to produce a gene product in one cell which is to be delivered to another location in the body. Secretion vectors may also facilitate the purification of desired proteins.

The present invention also relates to expression vectors capable of directing the expression of an inserted gene in a desired spatial or temporal manner or at a desired level. Such vectors may include sequences upstream of the 5' ESTs, such as promoters or upstream regulatory sequences.

Finally, the present invention may also be used for gene therapy to control or treat genetic diseases. Signal peptides may also be fused to heterologous proteins to direct their extracellular secretion.

Bacterial clones containing Bluescript plasmids having inserts containing the 5' ESTs of the present invention (SEQ ID NOs: 38-185 are presently stored at 80°C in 4% (v/v) glycerol in the inventor's laboratories under the designations listed next to the SEQ ID NOs in II). The inserts may be recovered from the deposited materials by growing the appropriate clones on a suitable medium. The Bluescript DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the EST insertion. The PCR product which corresponds to the 5' EST can then be manipulated using standard cloning techniques familiar to those skilled in the art.

One aspect of the present invention is a purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-185 or having a sequence complementary thereto. In one embodiment, the nucleic acid is recombinant.

5 Another aspect of the present invention is a purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-185 or one of the sequences complementary thereto.

Yet another aspect of the present invention is a purified or isolated nucleic acid comprising at least 15 consecutive bases of one of the sequences of SEQ ID NOs: 38-185 or one of the sequences complementary thereto. In one embodiment, the nucleic acid is  
10 recombinant.

A further aspect of the present invention is a purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-185 or one of the sequences complementary to the sequences of SEQ ID NOs: 38-185. In one embodiment, the nucleic acid is recombinant.

15 Another aspect of the present invention is a purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-185.

Still another aspect of the present invention is a method of making a cDNA encoding a human secretory protein, said human secretory protein being partially encoded by one of  
20 SEQ ID NOs 38-185, comprising the steps of contacting a collection of mRNA molecules from human cells with a primer comprising at least 15 consecutive nucleotides of a sequence complementary to one of SEQ ID NOs: 38-185; hybridizing said primer to an mRNA in said collection that encodes said protein; reverse transcribing said hybridized primer to make a first cDNA strand from said mRNA; making a second cDNA strand complementary to said first  
25 cDNA strand; and isolating the resulting cDNA encoding said protein comprising said first cDNA strand and said second cDNA strand.

Another aspect of the invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-185 or a fragment thereof of at least 10 amino acids, said cDNA being  
30 obtainable by the method described in the preceding paragraph. In one embodiment, the

cDNA comprises the full protein coding sequence of said protein which sequence is partially included in one of the sequences of SEQ ID NOs: 38-185.

Another aspect of the present invention is a method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-185, comprising  
5 the steps of obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-185; contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-185 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA; identifying a cDNA which hybridizes to said detectable probe; and isolating said cDNA which hybridizes to said probe.

10 Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-185 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the  
15 sequences of SEQ ID NOs: 38-185.

Another aspect of the present invention is a method of making a cDNA comprising one of the sequence of SEQ ID NOs: 38-185, comprising the steps of contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA; hybridizing said first primer to said polyA tail; reverse transcribing said  
20 mRNA to make a first cDNA strand; making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-185; and isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

Another aspect of the present invention is an isolated or purified cDNA encoding a  
25 human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-185 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-185.

30 In one embodiment of the method described in the two paragraphs above, the second cDNA strand is made by contacting said first cDNA strand with a first pair of primers, said



first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the sequences of SEQ ID NOs 38-185 and a third primer having a sequence therein which is included within the sequence of said first primer; performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product; contacting said first PCR product with a second pair of primers, said second pair of primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NOs: 38-185, and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and performing a second polymerase chain reaction, thereby generating a second PCR product.

One aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-185, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-185.

Another aspect of the present invention is the method described four paragraphs above in which the second cDNA strand is made by contacting said first cDNA strand with a second primer comprising at least 15 consecutive nucleotides of the sequences of SEQ ID NOs: 38-185; hybridizing said second primer to said first strand cDNA; and extending said hybridized second primer to generate said second cDNA strand.

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-185 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-185.

Another aspect of the present invention is a method of making a protein comprising one of the sequences of SEQ ID NOs: 186-333, comprising the steps of obtaining a cDNA encoding the full protein sequence partially included in one of the sequences of sequence of SEQ ID NOs: 38-185; inserting said cDNA in an expression vector such that said cDNA is

operably linked to a promoter; introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and isolating said protein.

Another aspect of the present invention is an isolated protein obtainable by the method described in the preceding paragraph.

5        Another aspect of the present invention is a method of obtaining a promoter DNA comprising the steps of obtaining DNAs located upstream of the nucleic acids of SEQ ID NOs: 38-185 or the sequences complementary thereto; screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and isolating said DNA comprising said identified promoter. In one embodiment, the obtaining step comprises  
10        chromosome walking from said nucleic acids of SEQ ID NOs: 38-185 or sequences complementary thereto. In another embodiment, the screening step comprises inserting said upstream sequences into a promoter reporter vector. In another embodiment, the screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

15        Another aspect of the present invention is an isolated promoter obtainable by the method described above.

Another aspect of the present invention is an isolated or purified protein comprising one of the sequences of SEQ ID NOs: 186-333.

20        Another aspect of the present invention is the inclusion of at least one of the sequences of SEQ ID NOs: 38-185, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-185, or a fragment thereof of at least 15 consecutive nucleotides in an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length. In one embodiment, the array includes at least two of the sequences of SEQ ID NOs: 38-185, the sequences complementary to the sequences of SEQ ID NOs: 38-185, or fragments thereof of  
25        at least 15 consecutive nucleotides. In another embodiment, the array includes at least five of the sequences of SEQ ID NOs: 38-185, the sequences complementary to the sequences of SEQ ID NOs: 38-185, or fragments thereof of at least 15 consecutive nucleotides.

Another aspect of the present invention is a promoter having a sequence selected from the group consisting of SEQ ID NOs: 31, 34, and 37.

30

### **Brief Description of the Drawings**

Figure 1 is a summary of a procedure for obtaining cDNAs which have been selected to include the 5' ends of the mRNAs from which they derived.

Figure 2 shows the distribution of Von Heijne scores for 5' ESTs in each of the categories described herein and the probability that these 5' ESTs encode a signal peptide.

Figure 3 summarizes a general method used to clone and sequence extended cDNAs containing sequences adjacent to 5' ESTs.

Figure 4 (description of promoters structure isolated from SignalTag 5' ESTs) provides a schematic description of promoters isolated and the way they are assembled with the corresponding 5' tags.

### **Detailed Description of the Preferred Embodiment**

Table IV is an analysis of the 43 amino acids located at the N terminus of all human SwissProt proteins to determine the frequency of false positives and false negatives using the techniques for signal peptide identification described herein.

Table V shows the distribution of 5' ESTs in each category described herein and the number of 5' ESTs in each category having a given minimum Von Heijne's score.

Table VI shows the distribution of 5' ESTs in each category described herein with respect to the tissue from which the 5' ESTs of the corresponding mRNA were obtained.

Table VII describes the transcription factor binding sites present in each of these promoters.

### **I. General Methods for Obtaining 5' ESTs derived from mRNAs with intact 5' ends**

In order to obtain the 5' ESTs of the present invention, mRNAs with intact 5' ends must be obtained. Currently, there are two approaches for obtaining such mRNAs with intact 5' ends as described below: either chemical (1) or enzymatic (2).

#### **1. Chemical Methods for Obtaining mRNAs having Intact 5' Ends**

One of these approaches is a chemical modification method involving derivatization of the 5' ends of the mRNAs and selection of the derivatized mRNAs. The 5' ends of eukaryotic mRNAs possess a structure referred to as a "cap" which comprises a guanosine

5 methylated at the 7 position. The cap is joined to the first transcribed base of the mRNA by a 5', 5'-triphosphate bond. In some instances, the 5' guanosine is methylated in both the 2 and 7 positions. Rarely, the 5' guanosine is trimethylated at the 2, 7 and 7 positions. In the chemical method for obtaining mRNAs having intact 5' ends, the 5' cap is specifically  
5 derivatized and coupled to a reactive group on an immobilizing substrate. This specific derivatization is based on the fact that only the ribose linked to the methylated guanosine at the 5' end of the mRNA and the ribose linked to the base at the 3' terminus of the mRNA, possess 2', 3'-cis diols.

10 Optionally, the 2', 3'-cis diol of the 3' terminal ribose may be chemically modified, substituted, converted, or eliminated, leaving only the ribose linked to the methylated guanosine at the 5' end of the mRNA with a 2', 3'-cis diol. A variety of techniques are available for eliminating the 2', 3'-cis diol on the 3' terminal ribose. For example, controlled alkaline hydrolysis may be used to generate mRNA fragments in which the 3' terminal ribose is a 3'-phosphate, 2'-phosphate or (2', 3')-cyclophosphate.  
15 Thereafter, the fragment which includes the original 3' ribose may be eliminated from the mixture through chromatography on an oligodT column. Alternatively, a base which lacks the 2', 3'-cis diol may be added to the 3' end of the mRNA using an RNA ligase such as T4 RNA ligase. Example 1 below describes a method for ligation of a nucleoside diphosphate to the 3' end of messenger RNA.

20

### EXAMPLE 1

#### Ligation of the Nucleoside Diphosphate pCp to the 3' End of mRNA

One  $\mu\text{g}$  of RNA was incubated in a final reaction medium of 10  $\mu\text{l}$  in the presence of 5 U of T<sub>4</sub> phage RNA ligase in the buffer provided by the manufacturer (Gibco -  
25 BRL), 40 U of the RNase inhibitor RNasin (Promega) and, 2  $\mu\text{l}$  of <sup>32</sup>pCp (Amersham #PB 10208). The incubation was performed at 37°C for 2 hours or overnight at 7-8°C.

Following modification or elimination of the 2', 3'-cis diol at the 3' ribose, the 2', 3'-cis diol present at the 5' end of the mRNA may be oxidized using reagents such as NaBH<sub>4</sub>,  
30 NaBH<sub>3</sub>CN, or sodium periodate, thereby converting the 2', 3'-cis diol to a dialdehyde.

Example 2 describes the oxidation of the 2', 3'-cis diol at the 5' end of the mRNA with sodium periodate.

## EXAMPLE 2

### 5      Oxidation of 2', 3'-cis diol at the 5' End of the mRNA with Sodium Periodate

0.1 OD unit of either a capped oligoribonucleotide of 47 nucleotides (including the cap) or an uncapped oligoribonucleotide of 46 nucleotides were treated as follows. The oligoribonucleotides were produced by *in vitro* transcription using the transcription kit "AmpliScribe T7" (Epicentre Technologies). As indicated below, the DNA template for the  
10      RNA transcript contained a single cytosine. To synthesize the uncapped RNA, all four NTPs were included in the *in vitro* transcription reaction. To obtain the capped RNA, GTP was replaced by an analogue of the cap, m7G(5')ppp(5')G. This compound, recognized by the polymerase, was incorporated into the 5' end of the nascent transcript during the initiation of transcription but was not incorporated during the extension step. Consequently, the resulting  
15      RNA contained a cap at its 5' end. The sequences of the oligoribonucleotides produced by the *in vitro* transcription reaction were:

+Cap:

5'm7GpppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-  
3' (SEQ ID NO:1)

20      -Cap:

5'-pppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3'  
(SEQ ID NO:2)

The oligoribonucleotides were dissolved in 9 µl of acetate buffer (0.1 M sodium acetate, pH 5.2) and 3 µl of freshly prepared 0.1 M sodium periodate solution. The mixture  
25      was incubated for 1 hour in the dark at 4°C or room temperature. Thereafter, the reaction was stopped by adding 4 µl of 10% ethylene glycol. The product was ethanol precipitated, resuspended in at least 10 µl of water or appropriate buffer and dialyzed against water.

The resulting aldehyde groups may then be coupled to molecules having a reactive  
30      amine group, such as hydrazine, carbazide, thiocarbazide or semicarbazide groups, in order to facilitate enrichment of the 5' ends of the mRNAs. Molecules having reactive amine groups

which are suitable for use in selecting mRNAs having intact 5' ends include avidin, proteins, antibodies, vitamins, ligands capable of specifically binding to receptor molecules, or oligonucleotides. Example 3 below describes the coupling of the resulting dialdehyde to biotin.

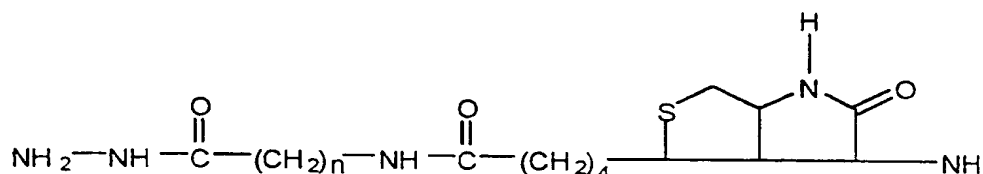
5

### EXAMPLE 3

#### Coupling of the Dialdehyde at the 5' End of Transcripts with Biotin

The oxidation product obtained in Example 2 was dissolved in 50  $\mu$ l of sodium acetate at a pH between 5 and 5.2 and 50  $\mu$ l of freshly prepared 0.02 M solution of biotin hydrazide in a methoxyethanol/water mixture (1:1) of formula:

10



In the compound used in these experiments,  $n=5$ . However, it will be appreciated that other commercially available hydrazides may also be used, such as molecules of the above formula in which  $n$  varies from 0 to 5. The mixture was then incubated for 2 hours at 37°C, precipitated with ethanol and dialyzed against distilled water. Example 4 demonstrates the specificity of the biotinylation reaction.

15

### EXAMPLE 4

20

#### Specificity of Biotinylation of Capped Transcripts

The specificity of the biotinylation for capped mRNAs was evaluated by gel electrophoresis of the following samples:

Sample 1. The 46 nucleotide uncapped *in vitro* transcript prepared as in Example 2 and labeled with  $^{32}$ Pcp as described in Example 1.

25

Sample 2. The 46 nucleotide uncapped *in vitro* transcript prepared as in Example 2, labeled with  $^{32}$ Pcp as described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of Example 3.

Sample 3. The 47 nucleotide capped *in vitro* transcript prepared as in Example 2 and labeled with  $^{32}\text{pCp}$  as described in Example 1.

Sample 4. The 47 nucleotide capped *in vitro* transcript prepared as in Example 2, labeled with  $^{32}\text{pCp}$  as described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of Example 3.

Samples 1 and 2 had identical migration rates, demonstrating that the uncapped RNAs were not oxidized and biotinylated. Sample 3 migrated more slowly than Samples 1 and 2, while Sample 4 exhibited the slowest migration. The difference in migration of the RNAs in Samples 3 and 4 demonstrates that the capped RNAs were specifically biotinylated.

10

In some cases, mRNAs having intact 5' ends may be enriched by binding the molecule containing a reactive amine group to a suitable solid phase substrate such as the inside of the vessel containing the mRNAs, magnetic beads, chromatography matrices, or nylon or nitrocellulose membranes. For example, where the molecule having a reactive amine group is biotin, the solid phase substrate may be coupled to avidin or streptavidin. Alternatively, where the molecule having the reactive amine group is an antibody or receptor ligand, the solid phase substrate may be coupled to the cognate antigen or receptor. Finally, where the molecule having a reactive amine group comprises an oligonucleotide, the solid phase substrate may comprise a complementary oligonucleotide.

The mRNAs having intact 5' ends may be released from the solid phase following the enrichment procedure. For example, where the dialdehyde is coupled to biotin hydrazide and the solid phase comprises streptavidin, the mRNAs may be released from the solid phase by simply heating to 95 degrees Celsius in 2% SDS. In some methods, the molecule having a reactive amine group may also be cleaved from the mRNAs having intact 5' ends following enrichment. Example 5 describes the capture of biotinylated mRNAs with streptavidin coated beads and the release of the biotinylated mRNAs from the beads following enrichment.

25

### EXAMPLE 5

#### Capture and Release of Biotinylated mRNAs Using Streptavidin Coated Beads

The streptavidin coated magnetic beads were prepared according to the manufacturer's instructions (CPG Inc., USA). The biotinylated mRNAs were added to a

30

hybridization buffer (1.5 M NaCl, pH 5 - 6). After incubating for 30 minutes, the unbound and nonbiotinylated material was removed. The beads were then washed several times in water with 1% SDS. The beads thus obtained were incubated for 15 minutes at 95°C in water containing 2% SDS.

- 5           Example 6 demonstrates the efficiency with which biotinylated mRNAs were recovered from the streptavidin coated beads.

### EXAMPLE 6

#### Efficiency of Recovery of Biotinylated mRNAs

- 10           The efficiency of the recovery procedure was evaluated as follows. Capped RNAs were labeled with <sup>32</sup>pCp, oxidized, biotinylated and bound to streptavidin coated beads as described above. Subsequently, the bound RNAs were incubated for 5, 15 or 30 minutes at 95°C in the presence of 2% SDS.

- 15           The products of the reaction were analyzed by electrophoresis on 12% polyacrylamide gels under denaturing conditions (7 M urea). The gels were subjected to autoradiography. During this manipulation, the hydrazone bonds were not reduced.

Increasing amounts of nucleic acids were recovered as incubation times in 2% SDS increased, demonstrating that biotinylated mRNAs were efficiently recovered.

- 20           In an alternative method for obtaining mRNAs having intact 5' ends, an oligonucleotide which has been derivatized to contain a reactive amine group is specifically coupled to mRNAs having an intact cap. Preferably, the 3' end of the mRNA is blocked prior to the step in which the aldehyde groups are joined to the derivatized oligonucleotide, as described above, so as to prevent the derivatized oligonucleotide from being joined to the 3' end of the mRNA. For example, pCp may be attached to the 3' end of the mRNA using T4 RNA ligase as described in example 1. However, as discussed above, blocking the 3' end of the mRNA is an optional step. Derivatized oligonucleotides may be prepared as described in Example 7.



**EXAMPLE 7**Derivatization of Oligonucleotides

An oligonucleotide phosphorylated at its 3' end was converted to a 3' hydrazide in 3' by treatment with an aqueous solution of hydrazine or of dihydrazide of the formula  
5  $H_2N(R1)NH_2$  at about 1 to 3 M, and at pH 4.5 at a temperature of 8°C overnight. This incubation was performed in the presence of a carbodiimide type agent soluble in water such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide at a final concentration of 0.3 M.

The derivatized oligonucleotide was then separated from the other agents and products using a standard technique for isolating oligonucleotides.

10 As discussed above, the mRNAs to be enriched may be treated to eliminate the 3' OH groups which may be present thereon. This may be accomplished by enzymatic ligation of sequences lacking a 3' OH, such as pCp, as described in Example 1. Alternatively, the 3' OH groups may be eliminated by alkaline hydrolysis as described in Example 8 below.

15

**EXAMPLE 8**Elimination of 3' OH Groups of mRNA Using Alkaline Hydrolysis

In a total volume of 100  $\mu$ l of 0.1 N sodium hydroxide, 1.5  $\mu$ g mRNA is incubated for 40 to 60 minutes at 4°C. The solution is neutralized with acetic acid and precipitated with ethanol.

20 Following the optional elimination of the 3' OH groups, the diol groups at the 5' ends of the mRNAs are oxidized as described below in Example 9.

**EXAMPLE 9**Oxidation of Diols of mRNA

25 Up to 1 OD unit of RNA was dissolved in 9  $\mu$ l of buffer (0.1 M sodium acetate, pH 6-7) or water and 3  $\mu$ l of freshly prepared 0.1 M sodium periodate solution. The reaction was incubated for 1 h in the dark at 4°C or room temperature. Following the incubation, the reaction was stopped by adding 4  $\mu$ l of 10% ethylene glycol. Thereafter the mixture was incubated at room temperature for 15 minutes. After ethanol precipitation, the product was  
30 resuspended in at least 10  $\mu$ l of water or appropriate buffer and dialyzed against water.

Following oxidation of the diol groups at the 5' ends of the mRNAs, the derivatized oligonucleotide was joined to the resulting aldehydes as described in Example 10.

### EXAMPLE 10

#### 5           Ligature of Aldehydes of mRNA to Derivatized Oligonucleotides

The oxidized mRNA was dissolved in an acidic medium such as 50  $\mu$ l of sodium acetate pH 4-6. Fifty  $\mu$ l of a solution of the derivatized oligonucleotide were added in order to obtain an mRNA:derivatized oligonucleotide ratio of 1:20. The mixture was reduced with a borohydride and incubated for 2 h at 37°C or overnight (14 h) at 10°C. The mixture was  
10 then ethanol precipitated, resuspended in 10  $\mu$ l or more of water or appropriate buffer and dialyzed against distilled water. If desired, the resulting product may be analyzed using acrylamide gel electrophoresis, HPLC analysis, or other conventional techniques.

Following the attachment of the derivatized oligonucleotide to the mRNAs, a reverse  
15 transcription reaction may be performed as described in Example 11 below.

### EXAMPLE 11

#### Reverse Transcription of mRNAs Ligatured to Derivatized Oligonucleotides

An oligodeoxyribonucleotide was derivatized as follows. Three OD units of an  
20 oligodeoxyribonucleotide of sequence 5'ATCAAGAATTCGCACGAGACCATTAA3' (SEQ ID NO:3) having 5'-OH and 3'-P ends were dissolved in 70  $\mu$ l of a 1.5 M hydroxybenzotriazole solution, pH 5.3, prepared in dimethylformamide/water (75:25) containing 2  $\mu$ g of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide. The mixture was incubated for 2 h 30 min at 22°C and then precipitated twice in LiClO<sub>4</sub>/acetone. The pellet  
25 was resuspended in 200  $\mu$ l of 0.25 M hydrazine and incubated at 8°C from 3 to 14 h. Following the hydrazine reaction, the mixture was precipitated twice in LiClO<sub>4</sub>/acetone.

The messenger RNAs to be reverse transcribed were extracted from blocks of placenta having sides of 2 cm which had been stored at -80°C. The total RNA was extracted using conventional acidic phenol techniques. Oligo-dT chromatography was used to purify  
30 the mRNAs. The integrity of the mRNAs was checked by Northern-blotting.

The diol groups on 7  $\mu$ g of the placental mRNAs were oxidized as described above in Example 9. The derivatized oligonucleotide was joined to the mRNAs as described in Example 10 above except that the precipitation step was replaced by an exclusion chromatography step to remove derivatized oligodeoxyribonucleotides which were not joined to mRNAs. Exclusion chromatography was performed as follows:

Ten ml of Ultrogel AcA34 (BioSeptra#230151) gel, a mix of agarose and acrylamide, were equilibrated in 50 ml of a solution of 10 mM Tris pH 8.0, 300 mM NaCl, 1 mM EDTA, and 0.05% SDS. The mixture was allowed to sediment. The supernatant was eliminated and the gel was resuspended in 50 ml of buffer. This procedure was repeated 2 or 3 times.

A glass bead (diameter 3 mm) was introduced into a 2 ml disposable pipette (length 25 cm). The pipette was filled with the gel suspension until the height of the gel stabilized at 1 cm from the top of the pipette. The column was then equilibrated with 20 ml of equilibration buffer (10 mM Tris HCl pH 7.4, 20 mM NaCl).

Ten  $\mu$ l of the mRNA which had reacted with the derivatized oligonucleotide were mixed in 39  $\mu$ l of 10 mM urea and 2  $\mu$ l of blue-glycerol buffer, which had been prepared by dissolving 5 mg of bromophenol blue in 60% glycerol (v/v), and passing the mixture through a 0.45  $\mu$ m diameter filter.

The column was then loaded with the mRNAs coupled to the oligonucleotide. As soon as the sample had penetrated, equilibration buffer was added. Hundred  $\mu$ l fractions were then collected. Derivatized oligonucleotide which had not been attached to mRNA appeared in fraction 16 and later fractions. Thus, fractions 3 to 15 were combined and precipitated with ethanol.

To determine whether the derivatized oligonucleotide was actually linked to mRNA, one tenth of the combined fractions were spotted twice on a nylon membrane and hybridized to a radioactive probe using conventional techniques. The  $^{32}$ P labeled probe used in these hybridizations was an oligodeoxyribonucleotide of sequence 5'TAATGGTCTCGTGCGAATTCTTGAT3' (SEQ ID NO:4) anticomplementary to the derivatized oligonucleotide. A signal observed after autoradiography, indicated that the derivatized oligonucleotide had been truly joined to the mRNA.

The remaining nine tenth of the mRNAs which had reacted with the derivatized oligonucleotide was reverse transcribed as follows. A reverse transcription reaction was

carried out with reverse transcriptase following the manufacturer's instructions and 50 pmol of nonamers with random sequence as primers.

To ensure that reverse transcription had been carried out through the cap structure, two types of experiments were performed.

5 In the first approach, after elimination of RNA of the cDNA:RNA heteroduplexes obtained from the reverse transcription reaction by an alkaline hydrolysis, a portion of the resulting single stranded cDNAs was spotted on a positively charged membrane and hybridized, using conventional methods, to a <sup>32</sup>P labeled probe having a sequence identical to that of the derivatized oligonucleotide. Control spots containing, 1 pmol, 100 fmol, 50 fmol,  
10 10 fmol and 1 fmol of a control oligodeoxyribonucleotide of sequence identical to that of the derivatized oligonucleotide were included. The signal observed in the spots containing the cDNA indicated that approximately 15 fmol of the derivatized oligonucleotide had been reverse transcribed. These results demonstrate that the reverse transcription can be performed through the cap and, in particular, that reverse transcriptase crosses the 5'-P-P-P-  
15 5' bond of the cap of eukaryotic messenger RNAs.

In the second type of experiment, the single stranded cDNAs obtained from the above first strand synthesis were used as template for PCR reactions. Two types of reactions were carried out. First, specific amplification of the mRNAs for alpha globin, dehydrogenase, pp15 and elongation factor E4 were carried out using the following pairs of  
20 oligodeoxyribonucleotide primers.

alpha-globin

GLO-S: 5'CCG ACA AGA CCA ACG TCA AGG CCG C3' (SEQ ID NO:5)

GLO-As: 5'TCA CCA GCA GGC AGT GGC TTA GGA G 3' (SEQ ID NO:6)

25

dehydrogenase

3 DH-S: 5'AGT GAT TCC TGC TAC TTT GGA TGG C3' (SEQ ID NO:7)

3 DH-As: 5'GCT TGG TCT TGT TCT GGA GTT TAG A3' (SEQ ID NO:8)

30

pp15

PP15-S: 5'TCC AGA ATG GGA GAC AAG CCA ATT T3' (SEQ ID NO:9)

PP15-As: 5'AGG GAG GAG GAA ACA GCG TGA GTC C3' (SEQ ID NO:10)

Elongation factor E4

EFA1-S: 5'ATG GGA AAG GAA AAG ACT CAT ATC A3' (SEQ ID NO:11)

5 EF1A-As: 5'AGC AGC AAC AAT CAG GAC AGC ACA G3' (SEQ ID NO:12)

Second, non specific amplifications were also carried out with the antisense oligodeoxyribonucleotides of the pairs described above and with a primer derived from the sequence of the derivatized oligodeoxyribonucleotide  
10 (5'ATCAAGAATTCGCACGAGACCATTAA3') (SEQ ID NO:13).

One twentieth of the following RT-PCR product samples were run on a 1.5% agarose gel and stained with ethidium bromide.

Sample 1: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the presence of cDNA.

15 Sample 2: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the absence of added cDNA.

Sample 3: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the presence of cDNA.

20 Sample 4: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the absence of added cDNA.

Sample 5: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the presence of cDNA.

Sample 6: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the absence of added cDNA.

25 Sample 7: The products of a PCR reaction using the EIF4 primers of SEQ ID NOs 11 and 12 in the presence of added cDNA.

Sample 8: The products of a PCR reaction using the EIF4 primers of SEQ ID NOs 11 and 12 in the absence of added cDNA.

30 A band of the size expected for the PCR product was observed only in samples 1, 3, 5 and 7, thus indicating the presence of the corresponding sequence in the cDNA population.

PCR reactions were also carried out with the antisense oligonucleotides of the globin and dehydrogenase primers (SEQ ID NOs 6 and 8) and an oligonucleotide whose sequence corresponds to that of the derivatized oligonucleotide. The presence of PCR products of the expected size in the samples equivalent to above samples 1 and 3 indicated that the derivatized oligonucleotide had been linked to mRNA.

The above examples summarize the chemical procedure for enriching mRNAs for those having intact 5' ends as illustrated in Figure 1. Further detail regarding the chemical approaches for obtaining such mRNAs are disclosed in International Application No. WO96/34981, published November 7, 1996, which is incorporated herein by reference. Strategies based on the above chemical modifications to the 5' cap structure may be utilized to generate cDNAs selected to include the 5' ends of the mRNAs from which they derived. In one version of such procedures, the 5' ends of the mRNAs are modified as described above. Thereafter, a reverse transcription reaction is conducted to extend a primer complementary to the 5' end of the mRNA. Single stranded RNAs are eliminated to obtain a population of cDNA/mRNA heteroduplexes in which the mRNA includes an intact 5' end. The resulting heteroduplexes may be captured on a solid phase coated with a molecule capable of interacting with the molecule used to derivatize the 5' end of the mRNA. Thereafter, the strands of the heteroduplexes are separated to recover single stranded first cDNA strands which include the 5' end of the mRNA. Second strand cDNA synthesis may then proceed using conventional techniques. For example, the procedures disclosed in WO 96/34981 or in Carninci. *et al.*, *Genomics* 37:327-336, 1996, the disclosures of which are incorporated herein by reference, may be employed to select cDNAs which include the sequence derived from the 5' end of the coding sequence of the mRNA.

Following ligation of the oligonucleotide tag to the 5' cap of the mRNA, a reverse transcription reaction is conducted to extend a primer complementary to the mRNA to the 5' end of the mRNA. Following elimination of the RNA component of the resulting heteroduplex using standard techniques, second strand cDNA synthesis is conducted with a primer complementary to the oligonucleotide tag.

## 2. Enzymatic Methods for Obtaining mRNAs having Intact 5' Ends

Other techniques for selecting cDNAs extending to the 5' end of the mRNA from which they are derived are fully enzymatic. Some versions of these techniques are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultes et perspectives nouvelles. Apports pour l'etude de la regulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EP0 625572 and Kato *et al.*, *Gene* 150:243-250, 1994, the disclosures of which are incorporated herein by reference.

Briefly, in such approaches, isolated mRNA is treated with alkaline phosphatase to remove the phosphate groups present on the 5' ends of uncapped incomplete mRNAs. Following this procedure, the cap present on full length mRNAs is enzymatically removed with a decapping enzyme such as T4 polynucleotide kinase or tobacco acid pyrophosphatase. An oligonucleotide, which may be either a DNA oligonucleotide or a DNA-RNA hybrid oligonucleotide having RNA at its 3' end, is then ligated to the phosphate present at the 5' end of the decapped mRNA using T4 RNA ligase. The oligonucleotide may include a restriction site to facilitate cloning of the cDNAs following their synthesis. Example 12 below describes one enzymatic method based on the doctoral thesis of Dumas.

### EXAMPLE 12

#### Enzymatic Approach for Obtaining 5' ESTs

Twenty micrograms of PolyA+ RNA were dephosphorylated using Calf Intestinal Phosphatase (Biolabs). After a phenol chloroform extraction, the cap structure of mRNA was hydrolysed using the Tobacco Acid Pyrophosphatase (purified as described by Shinshi *et al.*, *Biochemistry* 15: 2185-2190, 1976) and a hemi 5'DNA/RNA-3' oligonucleotide having an unphosphorylated 5' end, a stretch of adenosine ribophosphate at the 3' end, and an EcoRI site near the 5' end was ligated to the 5'P ends of mRNA using the T4 RNA ligase (Biolabs). Oligonucleotides suitable for use in this procedure are preferably 30 to 50 bases in length. Oligonucleotides having an unphosphorylated 5' end may be synthesized by adding a fluorochrome at the 5' end. The inclusion of a stretch of adenosine ribophosphates at the 3' end of the oligonucleotide increases ligation efficiency. It will be appreciated that the oligonucleotide may contain cloning sites other than EcoRI.

Following ligation of the oligonucleotide to the phosphate present at the 5' end of the decapped mRNA, first and second strand cDNA synthesis is carried out using conventional methods or those specified in EP0 625,572 and Kato *et al. supra*, and Dumas Milne Edwards, *supra*, the disclosures of which are incorporated herein by reference. The resulting cDNA may then be ligated into vectors such as those disclosed in Kato *et al. supra* or other nucleic acid vectors known to those skilled in the art using techniques such as those described in Sambrook *et al.*, Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbor Laboratory Press, 1989, the disclosure of which is incorporated herein by reference.

## II. Obtention and Characterization of the 5' ESTs of the Present Invention

The 5' ESTs of the present invention were obtained using the aforementioned chemical and enzymatic approaches for enriching mRNAs for those having intact 5' ends as described below.

### 1. Obtention of 5' ESTS Using mRNAs with Intact 5' Ends

First, mRNAs were prepared as described in Example 13 below.

#### EXAMPLE 13

##### Preparation of mRNA With Intact 5' Ends

Total human RNAs or polyA<sup>+</sup> RNAs derived from 29 different tissues were respectively purchased from LABIMO and CLONTECH and used to generate 44 cDNA libraries as follows. The purchased RNA had been isolated from cells or tissues using acid guanidium thiocyanate-phenol-chloroform extraction (Chomczynski and Sacchi, *Analytical Biochemistry* **162**:156-159, 1987). PolyA<sup>+</sup> RNA was isolated from total RNA (LABIMO) by two passes of oligo dT chromatography, as described by Aviv and Leder, *Proc. Natl. Acad. Sci. USA* **69**:1408-1412, 1972 in order to eliminate ribosomal RNA.

The quality and the integrity of the polyA<sup>+</sup> RNAs were checked. Northern blots hybridized with a globin probe were used to confirm that the mRNAs were not degraded.

Contamination of the polyA<sup>+</sup> mRNAs by ribosomal sequences was checked using Northern blots and a probe derived from the sequence of the 28S rRNA. Preparations of mRNAs with



less than 5% of rRNAs were used in library construction. To avoid constructing libraries with RNAs contaminated by exogenous sequences (prokaryotic or fungal), the presence of bacterial 16S ribosomal sequences or of two highly expressed fungal mRNAs was examined using PCR.

5        Following preparation of the mRNAs, the above described chemical and/or the enzymatic procedures for enriching mRNAs for those having intact 5' ends were employed to obtain 5' ESTs from various tissues. In both approaches, an oligonucleotide tag was attached to the 5' ends of the mRNAs. The oligonucleotide tag had an EcoRI site therein to facilitate later cloning procedures. To facilitate the processing of single stranded and double  
10        stranded cDNA obtained in the construction of the libraries, the same nucleotidic sequence was used to design the ligated oligonucleotide in both chemical and enzymatic approaches. Nevertheless, in the chemical procedure, the tag used was an oligodeoxyribonucleotide which was linked to the cap of the mRNA whereas in the enzymatic ligation, the tag was a chimeric  
15        hemi 5'DNA/RNA3' oligonucleotide which was ligated to the 5' end of decapped mRNA as described in example 12.

      Following attachment of the oligonucleotide tag to the mRNA by either the chemical or enzymatic methods, the integrity of the mRNA was examined by performing a Northern blot with 200 to 500 ng of mRNA using a probe complementary to the oligonucleotide tag before performing the first strand synthesis as described in example 14.

20

#### EXAMPLE 14

##### cDNA Synthesis Using mRNA Templates Having Intact 5' Ends

      For the mRNAs joined to oligonucleotide tags using both the chemical and enzymatic methods, first strand cDNA synthesis was performed using the Superscript II (Gibco BRL) or  
25        the Rnase H Minus M-MLV (Promega) reverse transcriptase with random nonamers as primers. In order to protect internal EcoRI sites in the cDNA from digestion at later steps in the procedure, methylated dCTP was used for first strand synthesis. After removal of RNA by an alkaline hydrolysis, the first strand of cDNA was precipitated using isopropanol in order to eliminate residual primers.

30        For both the chemical and the enzymatic methods, the second strand of the cDNA was synthesized with a Klenow fragment using a primer corresponding to the 5' end of the

ligated oligonucleotide described in Example 12. Preferably, the primer is 20-25 bases in length. Methylated dCTP was also used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following cDNA synthesis, the cDNAs were cloned into pBlueScript as described in  
5 Example 15 below.

### EXAMPLE 15

#### Cloning of cDNAs derived from mRNA with intact 5' ends into BlueScript

Following second strand synthesis, the ends of the cDNA were blunted with T4 DNA  
10 polymerase (Biolabs) and the cDNA was digested with EcoRI. Since methylated dCTP was used during cDNA synthesis, the EcoRI site present in the tag was the only hemi-methylated site, hence the only site susceptible to EcoRI digestion. The cDNA was then size fractionated using exclusion chromatography (AcA, Biosepra) and fractions corresponding to cDNAs of more than 150 bp were pooled and ethanol precipitated. The cDNA was directionally cloned  
15 into the SmaI and EcoRI ends of the phagemid pBlueScript vector (Stratagene). The ligation mixture was electroporated into bacteria and propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached were then selected as described in  
Example 16 below.

20

### EXAMPLE 16

#### Selection of Clones Having the Oligonucleotide Tag Attached Thereto

The plasmid DNAs containing 5' EST libraries made as described above were purified (Qiagen). A positive selection of the tagged clones was performed as follows.  
25 Briefly, in this selection procedure, the plasmid DNA was converted to single stranded DNA using gene II endonuclease of the phage F1 in combination with an exonuclease (Chang *et al.*, *Gene* 127:95-8, 1993) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA was then purified using paramagnetic beads as described by Fry *et al.*, *Biotechniques*, 13: 124-131, 1992. In this procedure, the single stranded DNA was  
30 hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide described in Example 13. Preferably, the primer has a length of 20-25

bases. Clones including a sequence complementary to the biotinylated oligonucleotide were captured by incubation with streptavidin coated magnetic beads followed by magnetic selection. After capture of the positive clones, the plasmid DNA was released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as the ThermoSequenase obtained from Amersham Pharmacia Biotech. Alternatively, protocols such as the one described in the Gene Trapper kit available from Gibco BRL may be used. The double stranded DNA was then electroporated into bacteria. The percentage of positive clones having the 5' tag oligonucleotide was estimated to typically rank between 90 and 98% using dot blot analysis.

Following electroporation, the libraries were ordered in 384-microtiter plates (MTP). A copy of the MTP was stored for future needs. Then the libraries were transferred into 96 MTP and sequenced as described below.

#### EXAMPLE 17

##### Sequencing of Inserts in Selected Clones

Plasmid inserts were first amplified by PCR on PE 9600 thermocyclers (Perkin-Elmer, Applied Biosystems Division, Foster City, CA), using standard SETA-A and SETA-B primers (Genset SA), AmpliTaqGold (Perkin-Elmer), dNTPs (Boehringer), buffer and cycling conditions as recommended by the Perkin-Elmer Corporation.

PCR products were then sequenced using automatic ABI Prism 377 sequencers (Perkin Elmer). Sequencing reactions were performed using PE 9600 thermocyclers with standard dye-primer chemistry and ThermoSequenase (Amersham Pharmacia Biotech). The primers used were either T7 or 21M13 (available from Genset SA) as appropriate. The primers were labeled with the JOE, FAM, ROX and TAMRA dyes. The dNTPs and ddNTPs used in the sequencing reactions were purchased from Boehringer. Sequencing buffer, reagent concentrations and cycling conditions were as recommended by Amersham.

Following the sequencing reaction, the samples were precipitated with ethanol, resuspended in formamide loading buffer, and loaded on a standard 4% acrylamide gel. Electrophoresis was performed for 2.5 hours at 3000V on an ABI 377 sequencer, and the sequence data were collected and analyzed using the ABI Prism DNA Sequencing Analysis Software, version 2.1.2.

## 2. Computer analysis of the Obtained 5' ESTs: Construction of NetGene and SignalTag databases

The sequence data from the 44 cDNA libraries made as described above were transferred to a proprietary database, where quality control and validation steps were performed. A proprietary base-caller, working using a Unix system, automatically flagged suspect peaks, taking into account the shape of the peaks, the inter-peak resolution, and the noise level. The proprietary base-caller also performed an automatic trimming. Any stretch of 25 or fewer bases having more than 4 suspect peaks was considered unreliable and was discarded. Sequences corresponding to cloning vector or ligation oligonucleotides were automatically removed from the EST sequences. However, the resulting EST sequences may contain 1 to 5 bases belonging to the above mentioned sequences at their 5' end. If needed, these can easily be removed on a case to case basis.

Following sequencing as described above, the sequences of the 5' ESTs were entered in NetGene™, a proprietary database called for storage and manipulation as described below. It will be appreciated by those skilled in the art that the data could be stored and manipulated on any medium which can be read and accessed by a computer. Computer readable media include magnetically, optically, or electronically readable media. For example, the computer readable media may be a hard disc, a floppy disc, a magnetic tape, CD-ROM, RAM, or ROM as well as other types of other media known to those skilled in the art.

In addition, the sequence data may be stored and manipulated in a variety of data processor programs in a diversity of formats. For instance, the sequence data may be stored as text in a word processing file, such as Microsoft WORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE.

The computer readable media on which the sequence information is stored may be in a personal computer, a network, a server or other computer systems known to those skilled in the art. The computer or other system preferably includes the storage media described above, and a processor for accessing and manipulating the sequence data. Once the sequence data has been stored, it may be manipulated and searched to locate those stored sequences which contain a desired nucleic acid sequence or which encode a protein having a particular functional domain. For example, the stored sequence information may be compared to other

known sequences to identify homologies, motifs implicated in biological function, or structural motifs.

Programs which may be used to search or compare the stored sequences include the MacPattern (EMBL), BLAST, and BLAST2 program series (NCBI), basic local alignment search tool programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (Altschul *et al*, *J. Mol. Biol.* **215**: 403, 1990) and FASTA (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* **85**: 2444, 1988). The BLAST programs then extend the alignments on the basis of defined match and mismatch criteria.

Motifs which may be detected using the above programs and those described in Example 28 include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

Before searching the cDNAs in the NetGene™ database for sequence motifs of interest, cDNAs derived from mRNAs which were not of interest were identified and eliminated from further consideration as described in Example 18 below.

### EXAMPLE 18

#### Elimination of Undesired Sequences from Further Consideration

5' ESTs in the NetGene™ database which were derived from undesired sequences such as transfer RNAs, ribosomal RNAs, mitochondrial RNAs, prokaryotic RNAs, fungal RNAs, Alu sequences, L1 sequences, or repeat sequences were identified using the FASTA and BLASTN programs with the parameters listed in Table I.

To eliminate 5' ESTs encoding tRNAs from further consideration, the 5' EST sequences were compared to the sequences of 1190 known tRNAs obtained from EMBL release 38, of which 100 were human. The comparison was performed using FASTA on both strands of the 5' ESTs. Sequences having more than 80% homology over more than 60 nucleotides were identified as tRNA. Of the 144,341 sequences screened, 26 were identified as tRNAs and eliminated from further consideration.

To eliminate 5' ESTs encoding rRNAs from further consideration, the 5' EST sequences were compared to the sequences of 2497 known rRNAs obtained from EMBL release 38, of which 73 were human. The comparison was performed using BLASTN on both strands of the 5' ESTs with the parameter S=108. Sequences having more than 80%  
5 homology over stretches longer than 40 nucleotides were identified as rRNAs. Of the 144,341 sequences screened, 3,312 were identified as rRNAs and eliminated from further consideration.

To eliminate 5' ESTs encoding mtRNAs from further consideration, the 5' EST sequences were compared to the sequences of the two known mitochondrial genomes for which the entire genomic sequences are available and all sequences transcribed from these  
10 mitochondrial genomes including tRNAs, rRNAs, and mRNAs for a total of 38 sequences. The comparison was performed using BLASTN on both strands of the 5' ESTs with the parameter S=108. Sequences having more than 80% homology over stretches longer than 40 nucleotides were identified as mtRNAs. Of the 144,341 sequences screened, 6,110 were  
15 identified as mtRNAs and eliminated from further consideration.

Sequences which might have resulted from exogenous contaminants were eliminated from further consideration by comparing the 5' EST sequences to release 46 of the EMBL bacterial and fungal divisions using BLASTN with the parameter S=144. All sequences having more than 90% homology over at least 40 nucleotides were identified as exogenous  
20 contaminants. Of the 42 cDNA libraries examined, the average percentages of prokaryotic and fungal sequences contained therein were 0.2% and 0.5% respectively. Among these sequences, only one could be identified as a sequence specific to fungi. The others were either fungal or prokaryotic sequences having homologies with vertebrate sequences or including repeat sequences which had not been masked during the electronic comparison.

25 In addition, the 5' ESTs were compared to 6093 Alu sequences and 1115 L1 sequences to mask 5' ESTs containing such repeat sequences. 5' ESTs including THE and MER repeats, SSTR sequences or satellite, micro-satellite, or telomeric repeats were also eliminated from further consideration. On average, 11.5% of the sequences in the libraries contained repeat sequences. Of this 11.5%, 7% contained Alu repeats, 3.3% contained L1  
30 repeats and the remaining 1.2% were derived from the other screened types of repetitive sequences. These percentages are consistent with those found in cDNA libraries prepared by

other groups. For example, the cDNA libraries of Adams *et al.* contained between 0% and 7.4% Alu repeats depending on the source of the RNA which was used to prepare the cDNA library (Adams *et al.*, *Nature* 377:174, 1996).

- 5           The sequences of those 5' ESTs remaining after the elimination of undesirable sequences were compared with the sequences of known human mRNAs to determine the accuracy of the sequencing procedures described above.

### EXAMPLE 19

10           Measurement of Sequencing Accuracy by Comparison to Known Sequences

To further determine the accuracy of the sequencing procedure described above, the sequences of 5' ESTs derived from known sequences were identified and compared to the original known sequences. First, a FASTA analysis with overhangs shorter than 5 bp on both ends was conducted on the 5' ESTs to identify those matching an entry in the public human mRNA database. The 6655 5' ESTs which matched a known human mRNA were then realigned with their cognate mRNA and dynamic programming was used to include substitutions, insertions, and deletions in the list of "errors" which would be recognized. Errors occurring in the last 10 bases of the 5' EST sequences were ignored to avoid the inclusion of spurious cloning sites in the analysis of sequencing accuracy.

- 20           This analysis revealed that the sequences incorporated in the NetGene™ database had an accuracy of more than 99.5%.

To determine the efficiency with which the above selection procedures select cDNAs which include the 5' ends of their corresponding mRNAs, the following analysis was performed.

25

### EXAMPLE 20

Determination of Efficiency of 5' EST Selection

- To determine the efficiency at which the above selection procedures isolated 5' ESTs which included sequences close to the 5' end of the mRNAs from which they derived, the sequences of the ends of the 5' ESTs derived from the elongation factor 1 subunit  $\alpha$  and

30

ferritin heavy chain genes were compared to the known cDNA sequences of these genes. Since the transcription start sites of both genes are well characterized, they may be used to determine the percentage of derived 5' ESTs which included the authentic transcription start sites.

5 For both genes, more than 95% of the obtained 5' ESTs actually included sequences close to or upstream of the 5' end of the corresponding mRNAs.

To extend the analysis of the reliability of the procedures for isolating 5' ESTs from ESTs in the NetGene™ database, a similar analysis was conducted using a database composed of human mRNA sequences extracted from GenBank database release 97 for  
10 comparison. The 5' ends of more than 85% of 5' ESTs derived from mRNAs included in the GeneBank database were located close to the 5' ends of the known sequence. As some of the mRNA sequences available in the GenBank database are deduced from genomic sequences, a 5' end matching with these sequences will be counted as an internal match. Thus, the method used here underestimates the yield of ESTs including the authentic 5' ends  
15 of their corresponding mRNAs.

The EST libraries made above included multiple 5' ESTs derived from the same mRNA. The sequences of such 5' ESTs were compared to one another and the longest 5' ESTs for each mRNA were identified. Overlapping cDNAs were assembled into continuous  
20 -sequences (contigs). The resulting continuous sequences were then compared to public databases to gauge their similarity to known sequences, as described in Example 21 below.

### EXAMPLE 21

#### Clustering of the 5' ESTs and Calculation of Novelty Indices for cDNA Libraries

25 For each sequenced EST library, the sequences were clustered by the 5' end. Each sequence in the library was compared to the others with BLASTN2 (direct strand, parameters S=107). ESTs with High Scoring Segment Pairs (HSPs) at least 25 bp long, having 95% identical bases and beginning closer than 10 bp from each EST 5' end were grouped. The longest sequence found in the cluster was used as representative of the group. A global  
30 clustering between libraries was then performed leading to the definition of super-contigs.



To assess the yield of new sequences within the EST libraries, a novelty rate (NR) was defined as:  $NR = 100 \times (\text{Number of new unique sequences found in the library} / \text{Total number of sequences from the library})$ . Typically, novelty rating ranged between 10% and 41% depending on the tissue from which the EST library was obtained. For most of the libraries, the random sequencing of 5' EST libraries was pursued until the novelty rate reached 20%.

Following characterization as described above, the collection of 5' ESTs in NetGene™ was screened to identify those 5' ESTs bearing potential signal sequences as described in Example 22 below.

## EXAMPLE 22

### Identification of Potential Signal Sequences in 5' ESTs

The 5' ESTs in the NetGene™ database were screened to identify those having an uninterrupted open reading frame (ORF) longer than 45 nucleotides beginning with an ATG codon and extending to the end of the EST. Approximately half of the cDNA sequences in NetGene™ contained such an ORF. The ORFs of these 5' ESTs were then searched to identify potential signal motifs using slight modifications of the procedures disclosed in Von Heijne, *Nucleic Acids Res.* 14:4683-4690, 1986, the disclosure of which is incorporated herein by reference. Those 5' EST sequences encoding a stretch of at least 15 amino acid long with a score of at least 3.5 in the Von Heijne signal peptide identification matrix were considered to possess a signal sequence. Those 5' ESTs which matched a known human mRNA or EST sequence and had a 5' end more than 20 nucleotides downstream of the known 5' end were excluded from further analysis. The remaining cDNAs having signal sequences therein were included in a database called SignalTag™.

To confirm the accuracy of the above method for identifying signal sequences, the analysis of Example 23 was performed.

**EXAMPLE 23**Confirmation of Accuracy of Identification of Potential Signal Sequences in 5' ESTs

The accuracy of the above procedure for identifying signal sequences encoding signal peptides was evaluated by applying the method to the 43 amino acids located at the N terminus of all human SwissProt proteins. The computed Von Heijne score for each protein was compared with the known characterization of the protein as being a secreted protein or a non-secreted protein. In this manner, the number of non-secreted proteins having a score higher than 3.5 (false positives) and the number of secreted proteins having a score lower than 3.5 (false negatives) could be calculated.

Using the results of the above analysis, the probability that a peptide encoded by the 5' region of the mRNA is in fact a genuine signal peptide based on its Von Heijne's score was calculated based on either the assumption that 10% of human proteins are secreted or the assumption that 20% of human proteins are secreted. The results of this analysis are shown in Figure 2 and in Table IV.

Using the above method of identification of secretory proteins, 5' ESTs of the following polypeptides known to be secreted were obtained: human glucagon, gamma interferon induced monokine precursor, secreted cyclophilin-like protein, human pleiotropin, and human biotinidase precursor. Thus, the above method successfully identified those 5' ESTs which encode a signal peptide.

To confirm that the signal peptide encoded by the 5' ESTs actually functions as a signal peptide, the signal sequences from the 5' ESTs may be cloned into a vector designed for the identification of signal peptides. Such vectors are designed to confer the ability to grow in selective medium only to host cells containing a vector with an operably linked signal sequence. For example, to confirm that a 5' EST encodes a genuine signal peptide, the signal sequence of the 5' EST may be inserted upstream and in frame with a non-secreted form of the yeast invertase gene in signal peptide selection vectors such as those described in U.S. Patent No. 5,536,637, the disclosure of which is incorporated herein by reference. Growth of host cells containing signal sequence selection vectors with the correctly inserted 5' EST signal sequence confirms that the 5' EST encodes a genuine signal peptide.

Alternatively, the presence of a signal peptide may be confirmed by cloning the extended cDNAs obtained using the ESTs into expression vectors such as pXT1 (as described below in example 30), or by constructing promoter-signal sequence-reporter gene vectors which encode fusion proteins between the signal peptide and an assayable reporter protein. After introduction of these vectors into a suitable host cell, such as COS cells or NIH 3T3 cells, the growth medium may be harvested and analyzed for the presence of the secreted protein. The medium from these cells is compared to the medium from control cells containing vectors lacking the signal sequence or extended cDNA insert to identify vectors which encode a functional signal peptide or an authentic secreted protein.

Those 5' ESTs which encoded a signal peptide, as determined by the method of Example 22 above, were further grouped into four categories based on their homology to known sequences as described in Example 24 below.

#### EXAMPLE 24

##### Categorization of 5' ESTs Encoding a Signal Peptide

Those 5' ESTs having a sequence not matching any known vertebrate sequence nor any publicly available EST sequence were designated "new." Of the sequences in the SignalTag™ database, 947 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those 5' ESTs having a sequence not matching any vertebrate sequence but matching a publicly known EST were designated "EST-ext", provided that the known EST sequence was extended by at least 40 nucleotides in the 5' direction. Of the sequences in the SignalTag™ database, 150 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those ESTs not matching any vertebrate sequence but matching a publicly known EST without extending the known EST by at least 40 nucleotides in the 5' direction were designated "EST." Of the sequences in the SignalTag™ database, 599 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those 5' ESTs matching a human mRNA sequence but extending the known sequence by at least 40 nucleotides in the 5' direction were designated "VERT-ext." Of the sequences in the SignalTag™ database, 23 of the 5' ESTs having a Von Heijne's score of at

least 3.5 fell into this category. Included in this category was a 5' EST which extended the known sequence of the human translocase mRNA by more than 200 bases in the 5' direction. A 5' EST which extended the sequence of a human tumor suppressor gene in the 5' direction was also identified.

5           Table V shows the distribution of 5' ESTs in each category and the number of 5' ESTs in each category having a given minimum von Heijne's score.

### 3. Evaluation of Spatial and Temporal Expression of mRNAs Corresponding to the 5'ESTs or Extended cDNAs

10

Each of the 5' ESTs was also categorized based on the tissue from which its corresponding mRNA was obtained, as described below in Example 25.

#### **EXAMPLE 25**

15

##### Categorization of Expression Patterns

Table VI shows the distribution of 5' ESTs in each of the above defined category with respect to the tissue from which the 5'ESTs of the corresponding mRNA were obtained.

20           Table II provides the sequence identification numbers of 5' EST sequences derived from different tissues, the categories in which these sequences fall, and the von Heijne's score of the signal peptides which they encode. The 5' EST sequences and the amino acid sequences they encode are provided in the appended sequence listings. Table III provides the sequence ID numbers of the 5' ESTs and the sequences of the signal peptides which they encode. The sequences of the 5' ESTs and the polypeptides they encode are provided in the sequence listing appended hereto.

25

The sequences of DNA SEQ ID NOs: 38-185 can readily be screened for any errors therein and any sequence ambiguities can be resolved by resequencing a fragment containing such errors or ambiguities on both strands. Such fragments may be obtained from the plasmids stored in the inventors' laboratory or can be isolated using the techniques described herein. Resolution of any such ambiguities or errors may be facilitated by using primers  
30           which hybridize to sequences located close to the ambiguous or erroneous sequences. For example, the primers may hybridize to sequences within 50-75 bases of the ambiguity or

error. Upon resolution of an error or ambiguity, the corresponding corrections can be made in the protein sequences encoded by the DNA containing the error or ambiguity.

In addition to categorizing the 5' ESTs with respect to their tissue of origin, the spatial and temporal expression patterns of the mRNAs corresponding to the 5' ESTs, as well as their expression levels, may be determined as described in Example 26 below. Characterization of the spatial and temporal expression patterns and expression levels of these mRNAs is useful for constructing expression vectors capable of producing a desired level of gene product in a desired spatial or temporal manner, as will be discussed in more detail below.

Furthermore, 5' ESTs whose corresponding mRNAs are associated with disease states may also be identified. For example, a particular disease may result from the lack of expression, over expression, or under expression of an mRNA corresponding to a 5' EST. By comparing mRNA expression patterns and quantities in samples taken from healthy individuals with those from individuals suffering from a particular disease, 5' ESTs responsible for the disease may be identified.

It will be appreciated that the results of the above characterization procedures for 5' ESTs also apply to extended cDNAs (obtainable as described below) which contain sequences adjacent to the 5' ESTs. It will also be appreciated that if desired, characterization may be delayed until extended cDNAs have been obtained rather than characterizing the ESTs themselves.

## EXAMPLE 26

### Evaluation of Expression Levels and Patterns of mRNAs

#### Corresponding to 5' ESTs or Extended cDNAs

Expression levels and patterns of mRNAs corresponding to 5' ESTs or extended cDNAs (obtainable as described below in example 27) may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277, the entire contents of which are hereby incorporated by reference. Briefly, a 5' EST, extended cDNA, or fragment thereof corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3,

T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the 5' EST or extended cDNA has 100 or more nucleotides. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (*i.e.* biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (*i.e.* RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

The 5' ESTs, extended cDNAs, or fragments thereof may also be tagged with nucleotide sequences for the serial analysis of gene expression (SAGE) as disclosed in UK Patent Application No. 2 305 241 A, the entire contents of which are incorporated by reference. In this method, cDNAs are prepared from a cell, tissue, organism or other source of nucleic acid for which gene expression patterns must be determined. The resulting cDNAs are separated into two pools. The cDNAs in each pool are cleaved with a first restriction endonuclease, called an anchoring enzyme, having a recognition site which is likely to be present at least once in most cDNAs. The fragments which contain the 5' or 3' most region of the cleaved cDNA are isolated by binding to a capture medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a so-called tagging endonuclease is ligated to the digested cDNAs in the first pool. Digestion with the second endonuclease produces short tag fragments from the cDNAs.

A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the tagging endonuclease to generate short tag fragments derived from the cDNAs in the second pool. The tags resulting from digestion of the first and second pools with the anchoring enzyme and the tagging endonuclease are ligated to one another to produce so-called ditags. In some embodiments, the ditags are concatamerized to produce ligation products containing from 2

to 200 ditags. The tag sequences are then determined and compared to the sequences of the 5' ESTs or extended cDNAs to determine which 5' ESTs or extended cDNAs are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of the 5' ESTs or extended cDNAs in the cell,  
5 tissue, organism, or other source of nucleic acids is obtained.

Quantitative analysis of gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of full length cDNAs (*i.e.* extended cDNAs which include the coding sequence for the signal peptide, the coding sequence for the mature protein, and a stop codon),  
10 extended cDNAs, 5' ESTs or fragments thereof of sufficient length to permit specific detection of gene expression. Preferably, the fragments are at least 15 nucleotides in length. More preferably, the fragments are at least 100 nucleotide long. More preferably, the fragments are more than 100 nucleotides in length. In some embodiments, the fragments may be more than 500 nucleotide long.

For example, quantitative analysis of gene expression may be performed with full length cDNAs as defined below, extended cDNAs, 5' ESTs, or fragments thereof in a complementary DNA microarray as described by Schena *et al.* (*Science* 270:467-470, 1995; *Proc. Natl. Acad. Sci. U.S.A.* 93:10614-10619, 1996). Full length cDNAs, extended cDNAs, 5' ESTs or fragments thereof are amplified by PCR and arrayed from 96-well microtiter  
20 plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm<sup>2</sup> microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a  
30 fluorescence laser scanning device fitted with a custom filter set. Accurate differential

expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of the expression of genes may also be performed with full length cDNAs, extended cDNAs, 5' ESTs, or fragments thereof in complementary DNA arrays as described by Pietu *et al.* (*Genome Research* 6:492-503, 1996). The full length cDNAs, extended cDNAs, 5' ESTs or fragments thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of the 5' ESTs or extended cDNAs can be done through high density nucleotide arrays as described by Lockhart *et al.* (*Nature Biotechnology* 14: 1675-1680, 1996) and Sosnowsky *et al.* (*Proc. Natl. Acad. Sci.* 94:1119-1123, 1997). Oligonucleotides of 15-50 nucleotides corresponding to sequences of the 5' ESTs or extended cDNAs are synthesized directly on the chip (Lockhart *et al.*, *supra*) or synthesized and then addressed to the chip (Sosnowsky *et al.*, *supra*). Preferably, the oligonucleotides are about 20 nucleotides in length.

cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart *et al.*, *supra* and application of different electric fields (Sonowsky *et al.*, *supra.*), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of the mRNA corresponding to the 5' EST or extended cDNA from which the oligonucleotide sequence has been designed.



### III. Use of 5' ESTs to Clone Extended cDNAs and to Clone the Corresponding Genomic DNAs

Once 5' ESTs which include the 5' end of the corresponding mRNAs have been selected using the procedures described above, they can be utilized to isolate extended cDNAs which contain sequences adjacent to the 5' ESTs. The extended cDNAs may include the entire coding sequence of the protein encoded by the corresponding mRNA, including the authentic translation start site, the signal sequence, and the sequence encoding the mature protein remaining after cleavage of the signal peptide. Such extended cDNAs are referred to herein as "full length cDNAs." Alternatively, the extended cDNAs may include only the sequence encoding the mature protein remaining after cleavage of the signal peptide, or only the sequence encoding the signal peptide.

Example 27 below describes a general method for obtaining extended cDNAs using 5' ESTs. Example 28 below provides experimental results, using the method explained in example 27, describing several extended cDNAs including the entire coding sequence and authentic 5' end of the corresponding mRNA for several secreted proteins.

The methods of Examples 27, 28, and 29 can also be used to obtain extended cDNAs which encode less than the entire coding sequence of the secreted proteins encoded by the genes corresponding to the 5' ESTs. In some embodiments, the extended cDNAs isolated using these methods encode at least 10 amino acids of one of the proteins encoded by the sequences of SEQ ID NOs: 38-185. In further embodiments, the extended cDNAs encode at least 20 amino acids of the proteins encoded by the sequences of SEQ ID NOs: 38-185. In further embodiments, the extended cDNAs encode at least 30 amino amino acids of the sequences of SEQ ID NOs: 38-185. In a preferred embodiment, the extended cDNAs encode a full length protein sequence, which includes the protein coding sequences of SEQ ID NOs: 38-185.

#### EXAMPLE 27

##### General Method for Using 5' ESTs to Clone and Sequence cDNAs which Include the Entire Coding Region and the Authentic 5' End of the Corresponding mRNA

The following general method has been used to quickly and efficiently isolate extended cDNAs having the authentic 5' ends of their corresponding mRNAs as well as

the full protein coding sequence and including sequence adjacent to the sequences of the 5' ESTs used to obtain them. This method may be applied to obtain extended cDNAs for any 5' EST in the NetGene™ database, including those 5' ESTs encoding polypeptides belonging to secreted proteins. The method is summarized in figure 3.

5

#### 1. Obtention of Extended cDNAs

##### *a) First strand synthesis*

The method takes advantage of the known 5' sequence of the mRNA. A reverse transcription reaction is conducted on purified mRNA with a poly 14dT primer containing a 49 nucleotide sequence at its 5' end allowing the addition of a known sequence at the end of the cDNA which corresponds to the 3' end of the mRNA. For example, the primer may have the following sequence: 5'-ATC GTT GAG ACT CGT ACC AGC AGA GTC ACG AGA GAG ACT ACA CGG TAC TGG TTT TTT TTT TTT TTVN -3' (SEQ ID NO:14). Those skilled in the art will appreciate that other sequences may also be added to the poly dT sequence and used to prime the first strand synthesis. Using this primer and a reverse transcriptase such as the Superscript II (Gibco BRL) or Rnase H Minus M-MLV (Promega) enzyme, a reverse transcript anchored at the 3' polyA site of the RNAs is generated.

15

20

After removal of the mRNA hybridized to the first cDNA strand by alkaline hydrolysis, the products of the alkaline hydrolysis and the residual poly dT primer are eliminated with an exclusion column such as an AcA34 (Biosepra) matrix as explained in Example 11.

##### *b) Second strand synthesis*

25

A pair of nested primers on each end is designed based on the known 5' sequence from the 5' EST and the known 3' end added by the poly dT primer used in the first strand synthesis. Softwares used to design primers are either based on GC content and melting temperatures of oligonucleotides, such as OSP (Illier and Green, *PCR Meth. Appl.* 1:124-128, 1991), or based on the octamer frequency disparity method (Griffais *et al.*, *Nucleic Acids Res.* 19: 3887-3891, 1991) such as PC-Rare (<http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html>).

Preferably, the nested primers at the 5' end are separated from one another by four to nine bases. The 5' primer sequences may be selected to have melting temperatures and specificities suitable for use in PCR.

Preferably, the nested primers at the 3' end are separated from one another by four to nine bases. For example, the nested 3' primers may have the following sequences: (5'- CCA GCA GAG TCA CGA GAG AGA CTA CAC GG -3'(SEQ ID NO:15), and 5'- CAC GAG AGA GAC TAC ACG GTA CTG G -3' (SEQ ID NO:16). These primers were selected because they have melting temperatures and specificities compatible with their use in PCR. However, those skilled in the art will appreciate that other sequences may also be used as primers.

The first PCR run of 25 cycles is performed using the Advantage Tth Polymerase Mix (Clontech) and the outer primer from each of the nested pairs. A second 20 cycle PCR using the same enzyme and the inner primer from each of the nested pairs is then performed on 1/2500 of the first PCR product. Thereafter, the primers and nucleotides are removed.

## 2. Sequencing of Full Length Extended cDNAs or Fragments Thereof

Due to the lack of position constraints on the design of 5' nested primers compatible for PCR use using the OSP software, amplicons of two types are obtained. Preferably, the second 5' primer is located upstream of the translation initiation codon thus yielding a nested PCR product containing the whole coding sequence. Such a full length extended cDNA undergoes a direct cloning procedure as described in section a. However, in some cases, the second 5' primer is located downstream of the translation initiation codon, thereby yielding a PCR product containing only part of the ORF. Such incomplete PCR products are submitted to a modified procedure described in section b.

### *a) Nested PCR products containing complete ORFs*

When the resulting nested PCR product contains the complete coding sequence, as predicted from the 5'EST sequence, it is cloned in an appropriate vector such as pED6dpc2, as described in section 3.

### *b) Nested PCR products containing incomplete ORFs*

When the amplicon does not contain the complete coding sequence, intermediate steps are necessary to obtain both the complete coding sequence and a PCR product containing the full coding sequence. The complete coding sequence can be assembled from several partial sequences determined directly from different PCR products as described in the following section.

Once the full coding sequence has been completely determined, new primers compatible for PCR use are designed to obtain amplicons containing the whole coding region. However, in such cases, 3' primers compatible for PCR use are located inside the 3' UTR of the corresponding mRNA, thus yielding amplicons which lack part of this region, *i.e.* the polyA tract and sometimes the polyadenylation signal, as illustrated in figure 3. Such full length extended cDNAs are then cloned into an appropriate vector as described in section 3.

*c) Sequencing extended cDNAs*

Sequencing of extended cDNAs is performed using a Die Terminator approach with the AmpliTaq DNA polymerase FS kit available from Perkin Elmer.

In order to sequence PCR fragments, primer walking is performed using software such as OSP to choose primers and automated computer software such as ASMG (Sutton *et al.*, *Genome Science Technol.* 1: 9-19, 1995) to construct contigs of walking sequences including the initial 5' tag using minimum overlaps of 32 nucleotides. Preferably, primer walking is performed until the sequences of full length cDNAs are obtained.

Completion of the sequencing of a given extended cDNA fragment is assessed as follows. Since sequences located after a polyA tract are difficult to determine precisely in the case of uncloned products, sequencing and primer walking processes for PCR products are interrupted when a polyA tract is identified in extended cDNAs obtained as described in case b. The sequence length is compared to the size of the nested PCR product obtained as described above. Due to the limited accuracy of the determination of the PCR product size by gel electrophoresis, a sequence is considered complete if the size of the obtained sequence is at least 70 % the size of the first nested PCR product. If the length of the sequence determined from the computer analysis is not at least 70 % of the length of the nested PCR product, these PCR products are cloned and the sequence of the insertion is determined. When Northern blot data are available, the size of the mRNA detected for a given PCR

product is used to finally assess that the sequence is complete. Sequences which do not fulfill the above criteria are discarded and will undergo a new isolation procedure.

Sequence data of all extended cDNAs are then transferred to a proprietary database, where quality controls and validation steps are carried out as described in  
5 example 15.

### 3. Cloning of Full Length Extended cDNAs

The PCR product containing the full coding sequence is then cloned in an appropriate vector. For example, the extended cDNAs can be cloned into the expression vector  
10 pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA) as follows. pED6dpc2 vector DNA is prepared with blunt ends by performing an EcoRI digestion followed by a fill in reaction. The blunt ended vector is dephosphorylated. After removal of PCR primers and ethanol precipitation, the PCR product containing the full coding sequence or the extended  
15 cDNA obtained as described above is phosphorylated with a kinase subsequently removed by phenol-Sevag extraction and precipitation. The double stranded extended cDNA is then ligated to the vector and the resulting expression plasmid introduced into appropriate host cells.

Since the PCR products obtained as described above are blunt ended molecules that can be cloned in either direction, the orientation of several clones for each PCR product is  
20 determined. Then, 4 to 10 clones are ordered in microtiter plates and subjected to a PCR reaction using a first primer located in the vector close to the cloning site and a second primer located in the portion of the extended cDNA corresponding to the 3' end of the mRNA. This second primer may be the antisense primer used in anchored PCR in the case of direct cloning (case a) or the antisense primer located inside the 3'UTR in the case of indirect cloning (case  
25 b). Clones in which the start codon of the extended cDNA is operably linked to the promoter in the vector so as to permit expression of the protein encoded by the extended cDNA are conserved and sequenced. In addition to the ends of cDNA inserts, approximately 50 bp of vector DNA on each side of the cDNA insert are also sequenced.

The cloned PCR products are then entirely sequenced according to the  
30 aforementioned procedure. In this case, contiguation of long fragments is then performed on walking sequences that have already contiguated for uncloned PCR products during

primer walking. Sequencing of cloned amplicons is complete when the resulting contigs include the whole coding region as well as overlapping sequences with vector DNA on both ends.

5     4. Computer analysis of Full Length Extended cDNA

Sequences of all full length extended cDNAs are then submitted to further analysis as described below. Before searching the extended full length cDNAs for sequences of interest, extended cDNAs which are not of interest (vector RNAs, transfer RNAs, ribosomal RNAs, mitochondrial RNAs, prokaryotic RNAs and fungal RNAs) are discarded using methods essentially similar to those described for 5'ESTs in Example 18.

10     *a) Identification of structural features*

Structural features, e.g. polyA tail and polyadenylation signal, of the sequences of full length extended cDNAs are subsequently determined as follows.

15     A polyA tail is defined as a homopolymeric stretch of at least 11 A with at most one alternative base within it. The polyA tail search is restricted to the last 100 nt of the sequence and limited to stretches of 11 consecutive A's because sequencing reactions are often not readable after such a polyA stretch. Stretches having more than 90% homology over 8 nucleotides are identified as polyA tails using BLAST2N.

20     To search for a polyadenylation signal, the polyA tail is clipped from the full-length sequence. The 50 bp preceding the polyA tail are first searched for the canonic polyadenylation AAUAAA signal and, if the canonic signal is not detected, for the alternative AUUAAA signal (Sheets *et al.*, *Nuc. Acids Res.* 18: 5799-5805, 1990). If neither of these consensus polyadenylation signals is found, the canonic motif is searched again allowing one mismatch to account for possible sequencing errors. More than 85 %  
25     of identified polyadenylation signals of either type actually ends 10 to 30 bp from the polyA tail. Alternative AUUAAA signals represents approximately 15 % of the total number of identified polyadenylation signals.

30     *b) Identification of functional features*

Functional features, e.g. ORFs and signal sequences, of the sequences of full length extended cDNAs were subsequently determined as follows.

The 3 upper strand frames of extended cDNAs are searched for ORFs defined as the maximum length fragments beginning with a translation initiation codon and ending with a stop codon. ORFs encoding at least 20 amino acids are preferred.

Each found ORF is then scanned for the presence of a signal peptide in the first  
5 50 amino-acids or, where appropriate, within shorter regions down to 20 amino acids or less in the ORF, using the matrix method of von Heijne (*Nuc. Acids Res.* **14**: 4683-4690, 1986), the disclosure of which is incorporated herein by reference as described in Example 22.

*c) Homology to either nucleotidic or proteic sequences*

10 Categorization of full-length sequences may be achieved using procedures essentially similar to those described for 5'ESTs in Example 24.

Extended cDNAs prepared as described above may be subsequently engineered to obtain nucleic acids which include desired portions of the extended cDNA using conventional  
15 techniques such as subcloning, PCR, or *in vitro* oligonucleotide synthesis. For example, nucleic acids which include only the full coding sequences (*i.e.* the sequences encoding the signal peptide and the mature protein remaining after the signal peptide is cleaved off) may be obtained using techniques known to those skilled in the art. Alternatively, conventional techniques may be applied to obtain nucleic acids which contain only the coding sequences  
20 for the mature protein remaining after the signal peptide is cleaved off or nucleic acids which contain only the coding sequences for the signal peptides.

Similarly, nucleic acids containing any other desired portion of the coding sequences for the secreted protein may be obtained. For example, the nucleic acid may contain at least 10 consecutive bases of an extended cDNA such as one of the extended cDNAs described  
25 below. In another embodiment, the nucleic acid may contain at least 15 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. Alternatively, the nucleic acid may contain at least 20 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. In another embodiment, the nucleic acid may contain at least 25 consecutive bases of an extended cDNA such as one of the extended cDNAs  
30 described below. In yet another embodiment, the nucleic acid may contain at least 40

consecutive bases of an extended cDNA such as one of the extended cDNAs described below.

Once an extended cDNA has been obtained, it can be sequenced to determine the amino acid sequence it encodes. Once the encoded amino acid sequence has been  
5 determined, one can create and identify any of the many conceivable cDNAs that will encode that protein by simply using the degeneracy of the genetic code. For example, allelic variants or other homologous nucleic acids can be identified as described below. Alternatively, nucleic acids encoding the desired amino acid sequence can be synthesized *in vitro*.

In a preferred embodiment, the coding sequence may be selected using the known  
10 codon or codon pair preferences for the host organism in which the cDNA is to be expressed.

The extended cDNAs derived from the 5' ESTS of the present invention were obtained as described in Example 28 below.

#### EXAMPLE 28

##### 15 Characterization of cloned extended cDNAs obtained using 5' ESTs

The procedure described in Example 27 above was used to obtain the extended cDNAs derived from the 5' ESTs of the present invention in a variety of tissues. The following list provides a few examples of thus obtained extended cDNAs.

Using this approach, the full length cDNA of SEQ ID NO:17 (internal identification  
20 number 48-19-3-G1-FL1) was obtained. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide MKKVLLLLITAILAVAVG (SEQ ID NO: 18) having a von Heijne score of 8.2.

The full length cDNA of SEQ ID NO:19 (internal identification number 58-34-2-E7-  
25 FL2) was also obtained using this procedure. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide MWWFQQGLSFLPSALVIWTSA (SEQ ID NO:20) having a von Heijne score of 5.5.

Another full length cDNA obtained using the procedure described above has the sequence of SEQ ID NO:21 (internal identification number 51-27-1-E8-FL1). This cDNA,  
30 falls into the "EST-ext" category described above and encodes the signal peptide MVLTTLPANSANSPPVNMPTTGPNLSYASSALSPCLT (SEQ ID NO:22) having a von Heijne score of 5.9.



The above procedure was also used to obtain a full length cDNA having the sequence of SEQ ID NO:23 (internal identification number 76-4-1-G5-FL1). This cDNA falls into the "EST-ext" category described above and encodes the signal peptide ILSTVTALTFAXA (SEQ ID NO:24) having a von Heijne score of 5.5.

5       The full length cDNA of SEQ ID NO:25 (internal identification number 51-3-3-B10-FL3) was also obtained using this procedure. This cDNA falls into the "new" category described above and encodes a signal peptide LVLTLCTLPLAVA (SEQ ID NO:26) having a von Heijne score of 10.1.

10       The full length cDNA of SEQ ID NO:27 (internal identification number 58-35-2-F10-FL2) was also obtained using this procedure. This cDNA falls into the "new" category described above and encodes a signal peptide LWLLFFLVTAIHA (SEQ ID NO:28) having a von Heijne score of 10.7.

15       Bacterial clones containing plasmids containing the full length cDNAs described above are presently stored in the inventor's laboratories under the internal identification numbers provided above. The inserts may be recovered from the stored materials by growing an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the cDNA insertion. The PCR product which corresponds to the cDNA can then be manipulated using standard cloning techniques familiar to those skilled in the art.

25       The polypeptides encoded by the extended cDNAs may be screened for the presence of known structural or functional motifs or for the presence of signatures, small amino acid sequences which are well conserved amongst the members of a protein family. The conserved regions have been used to derive consensus patterns or matrices included in the PROSITE data bank, in particular in the file prosite.dat (Release 13.0 of November 1995, located at <http://expasy.hcuge.ch/sprot/prosite.html>. Prosite\_convert and prosite\_scan

30

programs ([http://ulrec3.unil.ch/ftpserveur/prosite\\_scan](http://ulrec3.unil.ch/ftpserveur/prosite_scan)) may be used to find signatures on the extended cDNAs.

For each pattern obtained with the `prosite_convert` program from the `prosite.dat` file, the accuracy of the detection on a new protein sequence may be assessed by evaluating the frequency of irrelevant hits on the population of human secreted proteins included in the data bank SWISSPROT. The ratio between the number of hits on shuffled proteins (with a window size of 20 amino acids) and the number of hits on native (unshuffled) proteins may be used as an index. Every pattern for which the ratio is greater than 20% (one hit on shuffled proteins for 5 hits on native proteins) may be skipped during the search with `prosite_scan`. The program used to shuffle protein sequences (`db_shuffled`) and the program used to determine the statistics for each pattern in the protein data banks (`prosite_statistics`) are available on the ftp site [http://ulrec3.unil.ch/ftpserveur/prosite\\_scan](http://ulrec3.unil.ch/ftpserveur/prosite_scan).

In addition to PCR based methods for obtaining extended cDNAs, traditional hybridization based methods may also be employed. These methods may also be used to obtain the genomic DNAs which encode the mRNAs from which the 5' ESTs were derived, mRNAs corresponding to the extended cDNAs, or nucleic acids which are homologous to extended cDNAs or 5' ESTs. Example 29 below provides examples of such methods.

#### EXAMPLE 29

##### Methods for Obtaining cDNAs which include the Entire Coding Region and the Authentic 5' End of the Corresponding mRNA

A full length cDNA library can be made using the strategies described in Examples 13, 14, 15, and 16 above by replacing the random nonamer used in Example 14 with an oligo-dT primer. For instance, the oligonucleotide of SEQ ID NO:14 may be used.

Alternatively, a cDNA library or genomic DNA library may be obtained from a commercial source or made using techniques familiar to those skilled in the art. Such cDNA or genomic DNA libraries may be used to isolate extended cDNAs obtained from 5' EST or nucleic acids homologous to extended cDNAs or 5' EST as follows. The cDNA library or genomic DNA library is hybridized to a detectable probe comprising at least 10 consecutive nucleotides from the 5' EST or extended cDNA using conventional techniques. Preferably,

the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST or extended cDNA. More preferably, the probe comprises at least 20 to 30 consecutive nucleotides from the 5' EST or extended cDNA. In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST or extended cDNA.

5           Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual 2d Ed.*, Cold Spring Harbor Laboratory Press, 1989, the disclosure of which is incorporated herein by reference. The same techniques may be used to isolate genomic DNAs.

          Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are  
10 identified and isolated for further manipulation as follows. A probe comprising at least 10 consecutive nucleotides from the 5' EST or extended cDNA is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST or extended cDNA. More preferably, the probe comprises 20 to 30 consecutive nucleotides from the 5' EST or extended cDNA. In  
15 some embodiments, the probe comprises more than 30 nucleotides from the 5' EST or extended cDNA.

          Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, *in vitro* transcription, and non radioactive techniques.

          The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter  
20 and denatured. After blocking of non specific sites, the filter is incubated with the labeled probe for an amount of time sufficient to allow binding of the probe to cDNAs or genomic DNAs containing a sequence capable of hybridizing thereto.

          By varying the stringency of the hybridization conditions used to identify extended cDNAs or genomic DNAs which hybridize to the detectable probe, extended  
25 cDNAs having different levels of homology to the probe can be identified and isolated as described below.

1. Identification of Extended cDNA or Genomic cDNA Sequences Having a High Degree of Homology to the Labeled Probe

To identify extended cDNAs or genomic DNAs having a high degree of homology to the probe sequence, the melting temperature of the probe may be calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature ( $T_m$ ) is calculated using the formula:  $T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$  where  $N$  is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation  $T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G+C}) - (0.63\% \text{ formamide}) - (600/N)$  where  $N$  is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100  $\mu$ g denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100  $\mu$ g denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al.*, *supra*.

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to extended cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the  $T_m$ . For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 15-25°C below the  $T_m$ . Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

All of the foregoing hybridizations would be considered to be under "stringent" conditions.

Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization

temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

Extended cDNAs, nucleic acids homologous to extended cDNAs or 5' ESTs, or genomic DNAs which have hybridized to the probe are identified by autoradiography or other conventional techniques.

## 2. Obtention of Extended cDNA or Genomic cDNA Sequences Having Lower Degrees of Homology to the Labeled Probe

The above procedure may be modified to identify extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs having decreasing levels of homology to the probe sequence. For example, to obtain extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a sodium concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C.

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide.

Extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs which have hybridized to the probe are identified by autoradiography.

## 3. Determination of the Degree of Homology Between the Obtained Extended cDNAs and the Labeled Probe

If it is desired to obtain nucleic acids homologous to extended cDNAs, such as allelic variants thereof or nucleic acids encoding proteins related to the proteins encoded by the extended cDNAs, the level of homology between the hybridized nucleic acid and the

extended cDNA or 5' EST used as the probe may be further determined using BLAST2N; parameters may be adapted depending on the sequence length and degree of homology studied. To determine the level of homology between the hybridized nucleic acid and the extended cDNA or 5'EST from which the probe was derived, the nucleotide sequences of the hybridized nucleic acid and the extended cDNA or 5'EST from which the probe was derived are compared. For example, using the above methods, nucleic acids having at least 95% nucleic acid homology to the extended cDNA or 5'EST from which the probe was derived may be obtained and identified. Similarly, by using progressively less stringent hybridization conditions one can obtain and identify nucleic acids having at least 90%, at least 85%, at least 80% or at least 75% homology to the extended cDNA or 5'EST from which the probe was derived.

To determine whether a clone encodes a protein having a given amount of homology to the protein encoded by the extended cDNA or 5' EST, the amino acid sequence encoded by the extended cDNA or 5' EST is compared to the amino acid sequence encoded by the hybridizing nucleic acid. Homology is determined to exist when an amino acid sequence in the extended cDNA or 5' EST is closely related to an amino acid sequence in the hybridizing nucleic acid. A sequence is closely related when it is identical to that of the extended cDNA or 5' EST or when it contains one or more amino acid substitutions therein in which amino acids having similar characteristics have been substituted for one another. Using the above methods and algorithms such as FASTA with parameters depending on the sequence length and degree of homology studied, one can obtain nucleic acids encoding proteins having at least 95%, at least 90%, at least 85%, at least 80% or at least 75% homology to the proteins encoded by the extended cDNA or 5'EST from which the probe was derived.

In addition to the above described methods, other protocols are available to obtain extended cDNAs using 5' ESTs as outlined in the following paragraphs.

Extended cDNAs may be prepared by obtaining mRNA from the tissue, cell, or organism of interest using mRNA preparation procedures utilizing polyA selection procedures or other techniques known to those skilled in the art. A first primer capable of hybridizing to the polyA tail of the mRNA is hybridized to the mRNA and a reverse transcription reaction is performed to generate a first cDNA strand.

The first cDNA strand is hybridized to a second primer containing at least 10 consecutive nucleotides of the sequences of SEQ ID NOs 38-185. Preferably, the primer comprises at least 12, 15, or 17 consecutive nucleotides from the sequences of SEQ ID NOs 38-185. More preferably, the primer comprises 20 to 30 consecutive nucleotides from the sequences of SEQ ID NOs 38-185. In some embodiments, the primer comprises more than 30 nucleotides from the sequences of SEQ ID NOs 38-185. If it is desired to obtain extended cDNAs containing the full protein coding sequence, including the authentic translation initiation site, the second primer used contains sequences located upstream of the translation initiation site. The second primer is extended to generate a second cDNA strand complementary to the first cDNA strand. Alternatively, RT-PCR may be performed as described above using primers from both ends of the cDNA to be obtained.

Extended cDNAs containing 5' fragments of the mRNA may be prepared by hybridizing an mRNA comprising the sequence of the 5'EST for which an extended cDNA is desired with a primer comprising at least 10 consecutive nucleotides of the sequences complementary to the 5'EST and reverse transcribing the hybridized primer to make a first cDNA strand from the mRNAs. Preferably, the primer comprises at least 12, 15, or 17 consecutive nucleotides from the 5'EST. More preferably, the primer comprises 20 to 30 consecutive nucleotides from the 5'EST.

Thereafter, a second cDNA strand complementary to the first cDNA strand is synthesized. The second cDNA strand may be made by hybridizing a primer complementary to sequences in the first cDNA strand to the first cDNA strand and extending the primer to generate the second cDNA strand.

The double stranded extended cDNAs made using the methods described above are isolated and cloned. The extended cDNAs may be cloned into vectors such as plasmids or viral vectors capable of replicating in an appropriate host cell. For example, the host cell may be a bacterial, mammalian, avian, or insect cell.

Techniques for isolating mRNA, reverse transcribing a primer hybridized to mRNA to generate a first cDNA strand, extending a primer to make a second cDNA strand complementary to the first cDNA strand, isolating the double stranded cDNA and cloning the double stranded cDNA are well known to those skilled in the art and are described in *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc. 1997 and Sambrook *et al.*,

*Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989, the entire disclosures of which are incorporated herein by reference.

Alternatively, procedures such as the one described in Example 29 may be used for obtaining full length cDNAs or extended cDNAs. In this approach, full length or extended  
5 cDNAs are prepared from mRNA and cloned into double stranded phagemids as follows. The cDNA library in the double stranded phagemids is then rendered single stranded by treatment with an endonuclease, such as the Gene II product of the phage F1, and an exonuclease (Chang *et al.*, *Gene* 127:95-8, 1993). A biotinylated oligonucleotide comprising the sequence of a 5' EST, or a fragment containing at least 10 nucleotides thereof, is  
10 hybridized to the single stranded phagemids. Preferably, the fragment comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST. More preferably, the fragment comprises 20-30 consecutive nucleotides from the 5' EST. In some procedures, the fragment may comprise more than 30 consecutive nucleotides from the 5' EST.

Hybrids between the biotinylated oligonucleotide and phagemids having inserts  
15 containing the 5' EST sequence are isolated by incubating the hybrids with streptavidin coated paramagnetic beads and retrieving the beads with a magnet (Fry *et al.*, *Biotechniques*, 13: 124-131, 1992). Thereafter, the resulting phagemids containing the 5' EST sequence are released from the beads and converted into double stranded DNA using a primer specific for the 5' EST sequence. Alternatively, protocols such as the Gene Trapper kit (Gibco BRL)  
20 may be used. The resulting double stranded DNA is transformed into bacteria. Extended cDNAs containing the 5' EST sequence are identified by colony PCR or colony hybridization.

Using any of the above described methods in section III, a plurality of extended cDNAs containing full length protein coding sequences or sequences encoding only the  
25 mature protein remaining after the signal peptide is cleaved off may be provided as cDNA libraries for subsequent evaluation of the encoded proteins or use in diagnostic assays as described below.

#### IV. Expression of Proteins Encoded by Extended cDNAs Isolated Using 5' ESTs

30 Extended cDNAs containing the full protein coding sequences of their corresponding mRNAs or portions thereof, such as cDNAs encoding the mature protein, may be used to



express the encoded secreted proteins or portions thereof as described in Example 30 below. If desired, the extended cDNAs may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. It will be appreciated that a plurality of extended cDNAs containing the full protein coding sequences or portions thereof may be simultaneously cloned into expression vectors to create an expression library for analysis of the encoded proteins as described below.

### EXAMPLE 30

#### Expression of the Proteins Encoded by the Genes Corresponding to 5'ESTS or Portions Thereof

To express the proteins encoded by the genes corresponding to 5' ESTs (or portions thereof), full length cDNAs containing the entire protein coding region or extended cDNAs containing sequences adjacent to the 5' ESTs (or portions thereof) are obtained as described in Examples 27-29 and cloned into a suitable expression vector. If desired, the nucleic acids may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. The nucleic acids inserted into the expression vectors may also contain sequences upstream of the sequences encoding the signal peptide, such as sequences which regulate expression levels or sequences which confer tissue specific expression.

The nucleic acid encoding the protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The expression vector may be any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, *et al.*, U.S. Patent No. 5,082,767, incorporated herein by this reference.

The cDNA cloned into the expression vector may encode the entire protein (*i.e.* the signal peptide and the mature protein), the mature protein (*i.e.* the protein created by cleaving the signal peptide off), only the signal peptide or any other portion thereof.

The following is provided as one exemplary method to express the proteins encoded by the extended cDNAs corresponding to the 5' ESTs or the nucleic acids described above. First, the methionine initiation codon for the gene and the polyA signal of the gene are identified. If the nucleic acid encoding the polypeptide to be expressed lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the extended cDNA lacks a polyA signal, this sequence can be added to the construct by, for example, splicing out the polyA signal from pSG5 (Stratagene) using BglII and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the *gag* gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The extended cDNA or portion thereof encoding the polypeptide to be expressed is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the extended cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5' primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the extended cDNA is positioned with the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1 containing a poly A signal and prepared for this ligation (blunt/BglII).

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 µg/ml G418 (Sigma, St. Louis, Missouri). Preferably the expressed protein is released into the culture medium, thereby facilitating purification.

Alternatively, the extended cDNAs may be cloned into pED6dpc2 as described above. The resulting pED6dpc2 constructs may be transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded. Preferably, the protein expressed from the extended cDNA is released into the culture medium thereby facilitating purification.

Proteins in the culture medium are separated by gel electrophoresis. If desired, the proteins may be ammonium sulfate precipitated or separated based on size or charge prior to electrophoresis.

As a control, the expression vector lacking a cDNA insert is introduced into host cells  
5 or organisms and the proteins in the medium are harvested. The secreted proteins present in the medium are detected using techniques familiar to those skilled in the art such as Coomassie blue or silver staining or using antibodies against the protein encoded by the extended cDNA

Antibodies capable of specifically recognizing the protein of interest may be generated  
10 using synthetic 15-mer peptides having a sequence encoded by the appropriate 5' EST, extended cDNA, or portion thereof. The synthetic peptides are injected into mice to generate antibody to the polypeptide encoded by the 5' EST, extended cDNA, or portion thereof.

Secreted proteins from the host cells or organisms containing an expression vector which contains the extended cDNA derived from a 5' EST or a portion thereof are compared  
15 to those from the control cells or organism. The presence of a band in the medium from the cells containing the expression vector which is absent in the medium from the control cells indicates that the extended cDNA encodes a secreted protein. Generally, the band corresponding to the protein encoded by the extended cDNA will have a mobility near that expected based on the number of amino acids in the open reading frame of the extended  
20 cDNA. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, if the protein expressed from the above expression vectors does not contain sequences directing its secretion, the proteins expressed from host cells containing an expression vector with an insert encoding a secreted protein or portion thereof can be  
25 compared to the proteins expressed in control host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the desired protein or portion thereof is being expressed. Generally, the band will have the mobility expected for the secreted protein or portion  
30 thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

The protein encoded by the extended cDNA may be purified using standard immunochromatography techniques. In such procedures, a solution containing the secreted protein, such as the culture medium or a cell extract, is applied to a column having antibodies against the secreted protein attached to the chromatography matrix. The secreted protein is  
5 allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound secreted protein is then released from the column and recovered using standard techniques.

If antibody production is not possible, the extended cDNA sequence or portion thereof may be incorporated into expression vectors designed for use in purification schemes  
10 employing chimeric polypeptides. In such strategies, the coding sequence of the extended cDNA or portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera may be  $\beta$ -globin or a nickel binding polypeptide. A chromatography matrix having antibody to  $\beta$ -globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites may be engineered between the  $\beta$ -globin  
15 gene or the nickel binding polypeptide and the extended cDNA or portion thereof. Thus, the two polypeptides of the chimera may be separated from one another by protease digestion.

One useful expression vector for generating  $\beta$ -globin chimerics is pSG5 (Stratagene), which encodes rabbit  $\beta$ -globin. Intron II of the rabbit  $\beta$ -globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases  
20 the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, (*Basic Methods in Molecular Biology*, Davis, Dibner, and Battey, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using *in vitro*  
25 translation systems such as the *In vitro* Express<sup>TM</sup> Translation Kit (Stratagene).

Following expression and purification of the secreted proteins encoded by the 5' ESTs, extended cDNAs, or fragments thereof, the purified proteins may be tested for the ability to bind to the surface of various cell types as described in Example 31 below. It will be  
30 appreciated that a plurality of proteins expressed from these cDNAs may be included in a

panel of proteins to be simultaneously evaluated for the activities specifically described below, as well as other biological roles for which assays for determining activity are available.

### EXAMPLE 31

#### 5      Analysis of Secreted Proteins to Determine Whether they Bind to the Cell Surface

The proteins encoded by the 5' ESTs, extended cDNAs, or fragments thereof are cloned into expression vectors such as those described in Example 30. The proteins are purified by size, charge, immunochromatography or other techniques familiar to those skilled in the art. Following purification, the proteins are labeled using techniques known to those skilled in the art. The labeled proteins are incubated with cells or cell lines derived from a variety of organs or tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound protein. The labeled proteins are detected by autoradiography. Alternatively, unlabeled proteins may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various amounts of unlabeled protein are incubated along with the labeled protein. The amount of labeled protein bound to the cell surface decreases as the amount of competitive unlabeled protein increases. As a control, various amounts of an unlabeled protein unrelated to the labeled protein is included in some binding reactions. The amount of labeled protein bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated unlabeled protein, indicating that the protein encoded by the cDNA binds specifically to the cell surface.

As discussed above, secreted proteins have been shown to have a number of important physiological effects and, consequently, represent a valuable therapeutic resource. The secreted proteins encoded by the extended cDNAs or portions thereof made according to Examples 27-29 may be evaluated to determine their physiological activities as described below.

**EXAMPLE 32**

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Cytokine,  
Cell Proliferation or Cell Differentiation Activity

As discussed above, secreted proteins may act as cytokines or may affect cellular proliferation or differentiation. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein encoded by the extended cDNAs is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M<sup>+</sup> (preB M<sup>+</sup>), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c and CMK. The proteins encoded by the above extended cDNAs or portions thereof may be evaluated for their ability to regulate T cell or thymocyte proliferation in assays such as those described above or in the following references, which are incorporated herein by reference: *Current Protocols in Immunology*, Ed. by Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience; Takai *et al.* *J. Immunol.* **137**:3494-3500, 1986.; Bertagnolli *et al.*, *J. Immunol.* **145**:1706-1712, 1990.; Bertagnolli *et al.*, *Cell. Immunol.* **133**:327-341, 1991; Bertagnolli, *et al.*, *J. Immunol.* **149**:3778-3783, 1992; Bowman *et al.*, *J. Immunol.* **152**:1756-1761, 1994.

In addition, numerous assays for cytokine production and/or the proliferation of spleen cells, lymph node cells and thymocytes are known. These include the techniques disclosed in *Current Protocols in Immunology*, *supra* 1:3.12.1-3.12.14; and Schreiber In *Current Protocols in Immunology*, *supra* 1 : 6.8.1-6.8.8.

The proteins encoded by the cDNAs may also be assayed for the ability to regulate the proliferation and differentiation of hematopoietic or lymphopoietic cells. Many assays for such activity are familiar to those skilled in the art, including the assays in the following references, which are incorporated herein by reference: Bottomly *et al.*, In *Current Protocols in Immunology*, *supra* 1 : 6.3.1-6.3.12.; deVries *et al.*, *J. Exp. Med.* **173**:1205-1211, 1991; Moreau *et al.*, *Nature* **36**:690-692, 1988; Greenberger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**:2931-2938, 1983; Nordan, R., In *Current Protocols in Immunology*, *supra* 1 : 6.6.1-6.6.5; Smith *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**:1857-1861, 1986; Bennett *et al.*, in

*Current Protocols in Immunology supra* 1 : 6.15.1; Ciarletta *et al.*, In *Current Protocols in Immunology. supra* 1 : 6.13.1.

The proteins encoded by the cDNAs may also be assayed for their ability to regulate T-cell responses to antigens. Many assays for such activity are familiar to those skilled in the art, including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (*In Vitro* Assays for Mouse Lymphocyte Function), Chapter 6 (Cytokines and Their Cellular Receptors) and Chapter 7, (Immunologic Studies in Humans) in *Current Protocols in Immunology supra*; Weinberger *et al.*, *Proc. Natl. Acad. Sci. USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur. J. Immun.* 11:405-411, 1981; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988.

Those proteins which exhibit cytokine, cell proliferation, or cell differentiation activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which induction of cell proliferation or differentiation is beneficial. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

### EXAMPLE 33

#### Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Activity as Immune System Regulators

The proteins encoded by the cDNAs may also be evaluated for their effects as immune regulators. For example, the proteins may be evaluated for their activity to influence thymocyte or splenocyte cytotoxicity. Numerous assays for such activity are familiar to those skilled in the art including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (*In Vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic studies in Humans) in *Current Protocols in Immunology*, Coligan *et al.*, Eds, Greene Publishing Associates and Wiley-Interscience; Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann *et al.*, *J. Immunol.* 128:1968-1974, 1982; Handa *et al.*, *J. Immunol.* 135:1564-1572, 1985; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988;

Bowman *et al.*, *J. Virology* 61:1992-1998; Bertagnolli *et al.*, *Cell. Immunol.* 133:327-341, 1991; Brown *et al.*, *J. Immunol.* 153:3079-3092, 1994.

The proteins encoded by the cDNAs may also be evaluated for their effects on T-cell dependent immunoglobulin responses and isotype switching. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Maliszewski, *J. Immunol.* 144:3028-3033, 1990; Mond *et al.* in *Current Protocols in Immunology*, 1 : 3.8.1-3.8.16, *supra*.

The proteins encoded by the cDNAs may also be evaluated for their effect on immune effector cells, including their effect on Th1 cells and cytotoxic lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Chapter 3 (*In Vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic Studies in Humans) in *Current Protocols in Immunology*, *supra*; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Bertagnolli *et al.*, *J. Immunol.* 149:3778-3783, 1992.

The proteins encoded by the cDNAs may also be evaluated for their effect on dendritic cell mediated activation of naive T-cells. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Guery *et al.*, *J. Immunol.* 134:536-544, 1995; Inaba *et al.*, *J. Exp. Med.* 173:549-559, 1991; Macatonia *et al.*, *J. Immunol.* 154:5071-5079, 1995; Porgador *et al.*, *J. Exp. Med.* 182:255-260, 1995; Nair *et al.*, *J. Virol.* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.*, *J. Exp. Med.* 169:1255-1264, 1989; Bhardwaj *et al.*, *Journal of Clinical Investigation* 94:797-807, 1994; and Inaba *et al.*, *J. Exp. Med.* 172:631-640, 1990.

The proteins encoded by the cDNAs may also be evaluated for their influence on the lifetime of lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia* 7:659-670, 1993; Gorczyca *et al.*, *Cancer Res.* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *J. Immunol.* 145:4037-4045, 1990; Zamai *et al.*, *Cytometry* 14:891-897, 1993; Gorczyca *et al.*, *Int. J. Oncol.* 1:639-648, 1992.



The proteins encoded by the cDNAs may also be evaluated for their influence on early steps of T-cell commitment and development. Numerous assays for such activity are familiar to those skilled in the art, including without limitation the assays disclosed in the following references, which are incorporated herein by references: Antica *et al.*, *Blood* 84:111-117, 1994; Fine *et al.*, *Cell. Immunol.* 155:111-122, 1994; Galy *et al.*, *Blood* 85:2770-2778, 1995; Toki *et al.*, *Proc. Nat. Acad. Sci. USA* 88:7548-7551, 1991.

Those proteins which exhibit activity as immune system regulators activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of immune activity is beneficial. For example, the protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, *Leishmania* spp., *Plasmodium* and various fungal infections such as candidiasis. Of course, in this regard, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Alternatively, proteins encoded by extended cDNAs derived from the 5' ESTs of the present invention may be used in treatment of autoimmune disorders including, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention.

Using the proteins of the invention it may also be possible to regulate immune responses either up or down.

Down regulation may involve inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of  
5 activated T-cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active non-antigen-specific process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T  
10 cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after the end of exposure to the tolerizing agent. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions, such as, for example, B7 costimulation), e.g.,  
15 preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the  
20 transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation, can lead to the binding of the molecule to the natural  
25 ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this manner prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may  
30 avoid the necessity of repeated administration of these blocking reagents. To achieve

sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans.

5 Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, *Science* 257:789-792, 1992 and Turka *et al.*, *Proc. Natl. Acad. Sci USA*, 89:11102-11105, 1992. In addition, murine models of GVHD (see Paul ed., *Fundamental*  
10 *Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and  
15 autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor/ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which potentially involved in the disease process.  
20 Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/pr/pr mice or  
25 NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in OD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *supra*, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may involve either enhancing an existing immune response or eliciting an  
30 initial immune response as shown by the following examples. For instance, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases

of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory form of B lymphocyte antigens systemically.

Alternatively, antiviral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention or together with a stimulatory form of a soluble peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention and reintroducing the *in vitro* primed T cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to T cells *in vivo*, thereby activating the T cells.

In another application, upregulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules can be transfected with nucleic acids encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I  $\alpha$  chain and  $\beta_2$  microglobulin or an MHC class II  $\alpha$  chain and an MHC class II  $\beta$  chain to thereby express MHC class I or MHC class II proteins on the cell surface, respectively. Expression of the appropriate MHC class I or class II

molecules in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA  
5 encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject. Alternatively, as described in more detail below, genes encoding these immune system regulator proteins or nucleic acids regulating the expression of  
10 such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

#### EXAMPLE 34

##### Assaying the Proteins Expressed from Extended cDNAs 15 or Portions Thereof for Hematopoiesis Regulating Activity

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their hematopoiesis regulating activity. For example, the effect of the proteins on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following  
20 references, which are incorporated herein by reference: Johansson *et al.* *Cell. Biol.* 15:141-151, 1995; Keller *et al.*, *Mol. Cell. Biol.* 13:473-486, 1993; McClanahan *et al.*, *Blood* 81:2903-2915, 1993.

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their influence on the lifetime of stem cells and stem cell differentiation.  
25 Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Freshney, Methylcellulose Colony Forming Assays, in *Culture of Hematopoietic Cells*, Freshney, *et al.* Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama *et al.*, *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; McNiece and Briddell, in *Culture of Hematopoietic Cells*,  
30 *supra*; Neben *et al.*, *Exp. Hematol.* 22:353-359, 1994; Ploemacher and Cobblestone In

*Culture of Hematopoietic Cells, supra* 1-21, Spooncer *et al*, in *Culture of Hematopoietic Cells, supra* 163-179 and Sutherland in *Culture of Hematopoietic Cells, supra*. 139-162.

Those proteins which exhibit hematopoiesis regulatory activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoiesis is beneficial, such as in the treatment of myeloid or lymphoid cell deficiencies. Involvement in regulating hematopoiesis is indicated even by marginal biological activity in support of colony forming cells or of factor-dependent cell lines. For example, proteins supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, indicates utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells. Proteins supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (*i.e.*, traditional CSF activity) may be useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelosuppression. Proteins supporting the growth and proliferation of megakaryocytes and consequently of platelets allows prevention or treatment of various platelet disorders such as thrombocytopenia, and generally may be used in place of or complementary to platelet transfusions. Proteins supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells may therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in vivo* or *ex vivo* (*i.e.*, in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy. Alternatively, as described in more detail below, genes encoding hematopoiesis regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

**EXAMPLE 35**

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof  
for Regulation of Tissue Growth

5 The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their effect on tissue growth. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in International Patent Publication No. WO95/16035, International Patent Publication No. WO95/05846 and International Patent Publication No. WO91/07491, which are incorporated herein by reference.

10 Assays for wound healing activity include, without limitation, those described in: Winter, *Epidermal Wound Healing*, pps. 71-112, Maibach and Rovee, eds., Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, *J. Invest. Dermatol.* 71:382-84, 1978, which are incorporated herein by reference.

15 Those proteins which are involved in the regulation of tissue growth may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of tissue growth is beneficial. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

20 A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. *De novo* bone synthesis induced by an osteogenic agent  
25 contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of bone-  
30 forming cell progenitors. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or

by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein encoded by extended cDNAs derived from the 5' ESTs of the present invention is tendon/ligament formation. A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. *De novo* tendon/ligament-like tissue formation induced by a composition encoded by extended cDNAs derived from the 5' ESTs of the present invention contributes to the repair of tendon or ligaments defects of congenital, traumatic or other origin and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions encoded by extended cDNAs derived from the 5' ESTs of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.*, for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and



Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium) muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to generate. A protein of the invention may also exhibit angiogenic activity.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Alternatively, as described in more detail below, genes encoding tissue growth regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

### EXAMPLE 36

#### Assaying the Proteins Expressed from Extended cDNAs or Portions

#### Thereof for Regulation of Reproductive Hormones

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their ability to regulate reproductive hormones, such as follicle stimulating hormone. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference:

- 5 Vale *et al.*, *Endocrinol.* **91**:562-572, 1972; Ling *et al.*, *Nature* **321**:779-782, 1986; Vale *et al.*, *Nature* **321**:776-779, 1986; Mason *et al.*, *Nature* **318**:659-663, 1985; Forage *et al.*, *Proc. Natl. Acad. Sci. USA* **83**:3091-3095, 1986, Chapter 6.12 in *Current Protocols in Immunology*, Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Taub *et al.*, *J. Clin. Invest.* **95**:1370-1376, 1995; Lind *et al.*, *APMIS* **103**:140-146, 1995; Muller *et al.*, *Eur. J. Immunol.* **25**:1744-1748; Gruber *et al.*, *J. Immunol.* **152**:5860-5867, 1994;  
10 Johnston *et al.*, *J Immunol.* **153**:1762-1768, 1994.

Those proteins which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of reproductive hormones are beneficial. For example, a protein encoded by  
15 extended cDNAs derived from the 5' ESTs of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of FSH. Thus, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, alone or in heterodimers with a member of the inhibin  $\alpha$   
20 family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-B group, may be useful as a fertility inducing therapeutic, based upon the ability of  
25 activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885, the disclosure of which is incorporated herein by reference. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

30 Alternatively, as described in more detail below, genes encoding reproductive hormone regulating activity proteins or nucleic acids regulating the expression of such

proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

### EXAMPLE 37

5                   Assaying the Proteins Expressed from Extended cDNAs or  
                    Portions Thereof for Chemotactic/Chemokinetic Activity

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for chemotactic/chemokinetic activity. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may have chemotactic or  
10 chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as  
15 in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population.  
20 Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by  
25 the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include,  
30 without limitation, those described in: *Current Protocols in Immunology*, Ed by Coligan, Kruisbeek, Margulies, Shevach and Strober, Pub. Greene Publishing Associates and Wiley-

Interscience, Chapter 6.12: 6.12.1-6.12.28; Taub *et al.*, *J. Clin. Invest.* **95**:1370-1376, 1995; Lind *et al.*, *APMIS* **103**:140-146, 1995; Mueller *et al.*, *Eur. J. Immunol.* **25**:1744-1748; Gruber *et al.*, *J. Immunol.* **152**:5860-5867, 1994; Johnston *et al.* *J. Immunol.*, **153**:1762-1768, 1994.

5

### EXAMPLE 38

#### Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Blood Clotting

The proteins encoded by the extended cDNAs or portions thereof may also be  
10 evaluated for their effects on blood clotting. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Linet *et al.*, *J. Clin. Pharmacol.* **26**:131-140, 1986; Burdick *et al.*, *Thrombosis Res.* **45**:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* **5**:71-79, 1991; Schaub, *Prostaglandins* **35**:467-474, 1988.

15 Those proteins which are involved in the regulation of blood clotting may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of blood clotting is beneficial. For example, a protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulations disorders (including hereditary disorders, such as  
20 hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as infarction of cardiac and central nervous system vessels (e.g., stroke)). Alternatively, as described in more detail below, genes encoding blood  
25 clotting activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

### EXAMPLE 39

#### Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Involvement in Receptor/Ligand Interactions

30

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for their involvement in receptor/ligand interactions. Numerous assays for such involvement are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Chapter 7. 7.28.1-7.28.22 in *Current Protocols in Immunology*, Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Takai *et al.*, *Proc. Natl. Acad. Sci. USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160, 1989; Stoltenborg *et al.*, *J. Immunol. Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995; Gyuris *et al.*, *Cell* 75:791-803, 1993.

For example, the proteins encoded by extended cDNAs derived from the 5' ESTs of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions. Alternatively, as described in more detail below, genes encoding proteins involved in receptor/ligand interactions or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

#### EXAMPLE 40

##### Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Anti-Inflammatory Activity

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or

promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can  
5 be used to treat inflammatory conditions including chronic or acute conditions, including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine- or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over  
10 production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Alternatively, as described in more detail below, genes encoding anti-inflammatory activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

15

#### EXAMPLE 41

##### Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Tumor Inhibition Activity

The proteins encoded by the extended cDNAs or a portion thereof may also be  
20 evaluated for tumor inhibition activity. In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor  
25 growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth. Alternatively, as described in more detail below, genes tumor inhibition activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to  
30 increase or decrease the expression of the proteins as desired.

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein. Alternatively, as described in more detail below, genes encoding proteins involved in any of the above mentioned activities or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

## EXAMPLE 42

### Identification of Proteins which Interact with Polypeptides Encoded by Extended cDNAs

Proteins which interact with the polypeptides encoded by cDNAs derived from the 5' ESTs or fragments thereof, such as receptor proteins, may be identified using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the kit which is incorporated herein by reference,

the the cDNAs derived from 5' ESTs, or fragments thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. cDNAs in a cDNA library which encode proteins which might interact with the polypeptides encoded by the extended cDNAs or portions thereof are  
5 inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4  
10 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain plasmids encoding proteins which interact with the polypeptide encoded by the extended cDNAs or portions thereof.

Alternatively, the system described in Lustig *et al.*, *Methods in Enzymology* 283: 83-99, 1997, and in U.S. Patent No. 5,654,150, the disclosure of which is incorporated herein by  
15 reference, may be used for identifying molecules which interact with the polypeptides encoded by extended cDNAs. In such systems, *in vitro* transcription reactions are performed on a pool of vectors containing extended cDNA inserts cloned downstream of a promoter which drives *in vitro* transcription. The resulting pools of mRNAs are introduced into *Xenopus laevis* oocytes. The oocytes are then assayed for a desired activity.

20 Alternatively, the pooled *in vitro* transcription products produced as described above may be translated *in vitro*. The pooled *in vitro* translation products can be assayed for a desired activity or for interaction with a known polypeptide.

Proteins or other molecules interacting with polypeptides encoded by extended cDNAs can be found by a variety of additional techniques. In one method, affinity  
25 columns containing the polypeptide encoded by the extended cDNA or a portion thereof can be constructed. In some versions, of this method the affinity column contains chimeric proteins in which the protein encoded by the extended cDNA or a portion thereof is fused to glutathione S-transferase. A mixture of cellular proteins or pool of expressed proteins as described above and is applied to the affinity column. Proteins  
30 interacting with the polypeptide attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen *et al.*, *Electrophoresis* 18:588-598,



1997, the disclosure of which is incorporated herein by reference. Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

5 Proteins interacting with polypeptides encoded by extended cDNAs or portions thereof can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow, *Analytical Biochemistry* 246:1-6, 1997, the disclosure of which is incorporated herein by reference. The main advantage of the method is that it allows the determination of the association rate between the protein and other interacting  
10 molecules. Thus, it is possible to specifically select interacting molecules with a high or low association rate. Typically a target molecule is linked to the sensor surface (through a carboxymethyl dextran matrix) and a sample of test molecules is placed in contact with the target molecules. The binding of a test molecule to the target molecule causes a change in the refractive index and/ or thickness. This change is detected by the  
15 Biosensor provided it occurs in the evanescent field (which extend a few hundred nanometers from the sensor surface). In these screening assays, the target molecule can be one of the polypeptides encoded by extended cDNAs or a portion thereof and the test sample can be a collection of proteins extracted from tissues or cells, a pool of expressed proteins, combinatorial peptide and/ or chemical libraries, or phage displayed peptides.  
20 The tissues or cells from which the test proteins are extracted can originate from any species.

In other methods, a target protein is immobilized and the test population is a collection of unique polypeptides encoded by the extended cDNAs or portions thereof.

To study the interaction of the proteins encoded by the extended cDNAs or  
25 portions thereof with drugs, the microdialysis coupled to HPLC method described by Wang *et al.*, *Chromatographia* 44:205-208, 1997 or the affinity capillary electrophoresis method described by Busch *et al.*, *J. Chromatogr.* 777:311-328, 1997, the disclosures of which are incorporated herein by reference can be used.

30 It will be appreciated by those skilled in the art that the proteins expressed from the extended cDNAs or portions may be assayed for numerous activities in addition to those

specifically enumerated above. For example, the expressed proteins may be evaluated for applications involving control and regulation of inflammation, tumor proliferation or metastasis, infection, or other clinical conditions. In addition, the proteins expressed from the extended cDNAs or portions thereof may be useful as nutritional agents or cosmetic agents.

5       The proteins expressed from the cDNAs or portions thereof may be used to generate antibodies capable of specifically binding to the expressed protein or fragments thereof as described in Example 40 below. The antibodies may be capable of binding a full length protein encoded by a cDNA derived from a 5' EST, a mature protein (*i.e.* the protein generated by cleavage of the signal peptide) encoded by a cDNA derived from a 5' EST, or a signal  
10    peptide encoded by a cDNA derived from a 5' EST. Alternatively, the antibodies may be capable of binding fragments of at least 10 amino acids of the proteins encoded by the above cDNAs. In some embodiments, the antibodies may be capable of binding fragments of at least 15 amino acids of the proteins encoded by the above cDNAs. In other embodiments, the antibodies may be capable of binding fragments of at least 25 amino acids of the proteins  
15    expressed from the extended cDNAs which comprise at least 25 amino acids of the proteins encoded by the above cDNAs. In further embodiments, the antibodies may be capable of binding fragments of at least 40 amino acids of the proteins encoded by the above cDNAs.

### EXAMPLE 43

#### 20       Production of an Antibody to a Human Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells as described in Example 30. The concentration of protein in the ~~final~~ preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few  $\mu\text{g/ml}$ . Monoclonal or polyclonal antibody to the protein can then be  
25    prepared as follows:

#### 1. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of  
30    Kohler, and Milstein, *Nature* 256:495, 1975 or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein or

peptides derived therefrom over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, *Meth. Enzymol.* 70:419, 1980, the disclosure of which is incorporated herein by reference and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis *et al.* in *Basic Methods in Molecular Biology* Elsevier, New York. Section 21-2, the disclosure of which is incorporated herein by reference.

15

## 2. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein or peptides derived therefrom, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals response vary depending on site of inoculations and doses, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, *et al.*, *J. Clin. Endocrinol. Metab.* 33:988-991 (1971), the disclosure of which is incorporated herein by reference.

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, *et al.*, Chap. 19 in: *Handbook of Experimental Immunology* D. Wier

30

(ed) Blackwell (1973) , the disclosure of which is incorporated herein by reference. Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12  $\mu$ M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980) , the disclosure of which is incorporated herein by reference..

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

#### V. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof as Reagents

The 5' ESTs of the present invention (or cDNAs or genomic DNAs obtainable therefrom) may be used as reagents in isolation procedures, diagnostic assays, and forensic procedures. For example, sequences from the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be detectably labeled and used as probes to isolate other sequences capable of hybridizing to them. In addition, sequences from 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be used to design PCR primers to be used in isolation, diagnostic, or forensic procedures.

##### 1. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof in Isolation, Diagnostic and Forensic Procedures

#### EXAMPLE 44

##### Preparation of PCR Primers and Amplification of DNA

The 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) may be used to prepare PCR primers for a variety of applications, including isolation procedures for cloning nucleic acids capable of hybridizing to such sequences, diagnostic techniques and

forensic techniques. The PCR primers are at least 10 bases, and preferably at least 12, 15, or 17 bases in length. More preferably, the PCR primers are at least 20-30 bases in length. In some embodiments, the PCR primers may be more than 30 bases in length. It is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering, White Ed. in *Methods in Molecular Biology* 67: Humana Press, Totowa 1997, the disclosure of which is incorporated herein by reference. In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites.

#### EXAMPLE 45

##### Use of 5'ESTs as Probes

Probes derived from 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom), including full length cDNAs or genomic sequences, may be labeled with detectable labels familiar to those skilled in the art, including radioisotopes and non-radioactive labels, to provide a detectable probe. The detectable probe may be single stranded or double stranded and may be made using techniques known in the art, including *in vitro* transcription, nick translation, or kinase reactions. A nucleic acid sample containing a sequence capable of hybridizing to the labeled probe is contacted with the labeled probe. If the nucleic acid in the sample is double stranded, it may be denatured prior to contacting the probe. In some applications, the nucleic acid sample may be immobilized on a surface such as a nitrocellulose or nylon membrane. The nucleic acid sample may comprise nucleic acids obtained from a variety of sources, including genomic DNA, cDNA libraries, RNA, or tissue samples.

Procedures used to detect the presence of nucleic acids capable of hybridizing to the detectable probe include well known techniques such as Southern blotting, Northern blotting,

dot blotting, colony hybridization, and plaque hybridization. In some applications, the nucleic acid capable of hybridizing to the labeled probe may be cloned into vectors such as expression vectors, sequencing vectors, or *in vitro* transcription vectors to facilitate the characterization and expression of the hybridizing nucleic acids in the sample. For example, such techniques  
5 may be used to isolate and clone sequences in a genomic library or cDNA library which are capable of hybridizing to the detectable probe as described in Example 30 above.

PCR primers made as described in Example 44 above may be used in forensic analyses, such as the DNA fingerprinting techniques described in Examples 46-50 below. Such analyses may utilize detectable probes or primers based on the sequences of the the 5'  
10 ESTs or of cDNAs or genomic DNAs isolated using the 5' ESTs.

#### EXAMPLE 46

##### Forensic Matching by DNA Sequencing

In one exemplary method, DNA samples are isolated from forensic specimens of, for  
15 example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers based on a number of the 5' ESTs of Example 25, or cDNAs or genomic DNAs isolated therefrom as described above, is then utilized in accordance with Example 44 to amplify DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a test subject. Each of these identification DNAs is then  
20 sequenced using standard techniques, and a simple database comparison determines the differences, if any, between the sequences from the subject and those from the sample. Statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large  
25 number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences of 100 bases in length are used to prove identity between the suspect and the sample.

**EXAMPLE 47****Positive Identification by DNA Sequencing**

The technique outlined in the previous example may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of 5'EST sequences from Example 25, or cDNA or genomic DNA sequences obtainable therefrom. Preferably, 20 to 50 different primers are used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question in accordance with Example 44. Each of these DNA segments is sequenced, using the methods set forth in Example 46. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological specimen with that individual.

**EXAMPLE 48****Southern Blot Forensic Identification**

The procedure of Example 47 is repeated to obtain a panel of at least 10 amplified sequences from an individual and a specimen. Preferably, the panel contains at least 50 amplified sequences. More preferably, the panel contains 100 amplified sequences. In some embodiments, the panel contains 200 amplified sequences. This PCR-generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern blotting see Davis *et al.* (Basic Methods in Molecular Biology, 1986, Elsevier Press. pp 62-65), the disclosure of which is incorporated herein by reference..

A panel of probes based on the sequences of 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom), or fragments thereof of at least 10 bases, are radioactively or colorimetrically labeled using methods known in the art, such as nick translation or end labeling, and hybridized to the Southern blot using techniques known in the art (Davis *et al.*, supra). Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from

the 5' EST (or cDNAs or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the 5' EST (or cDNAs or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST (or cDNAs or genomic DNAs obtainable therefrom).

5 Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of 5' EST (or cDNAs or genomic DNAs obtainable therefrom) will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the  
10 number of 5' EST (or cDNAs or genomic DNAs obtainable therefrom) probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

#### EXAMPLE 49

##### 15 Dot Blot Identification Procedure

Another technique for identifying individuals using the 5' EST sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length are synthesized that correspond to at least 10,  
20 preferably 50 sequences from the 5' ESTs or cDNAs or genomic DNAs obtainable therefrom. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with P<sup>32</sup> using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose  
25 filter containing the genomic sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis *et al.*, *supra*). The <sup>32</sup>P labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA. Tetramethylammonium chloride is useful for identifying clones containing small numbers of  
30 nucleotide mismatches (Wood *et al.*, *Proc. Natl. Acad. Sci. USA* 82(6):1585-1588, 1985)



which is hereby incorporated by reference. A unique pattern of dots distinguishes one individual from another individual.

5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) or oligonucleotides containing at least 10 consecutive bases from these sequences can be used as probes in the following alternative fingerprinting technique. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom).

Preferably, a plurality of probes having sequences from different genes are used in the alternative fingerprinting technique. Example 50 below provides a representative alternative fingerprinting procedure in which the probes are derived from 5'EST.

15

### EXAMPLE 50

#### Alternative "Fingerprint" Identification Technique

20-mer oligonucleotides are prepared from a large number, e.g. 50, 100, or 200, of 5'EST using commercially available oligonucleotide services such as Genset, Paris, France. Cell samples from the test subject are processed for DNA using techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes such as EcoRI and XbaI. Following digestion, samples are applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred onto nitrocellulose using standard Southern blotting techniques.

25

10 ng of each of the oligonucleotides are pooled and end-labeled with  $^{32}\text{P}$ . The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

30

It is additionally contemplated within this example that the number of probe sequences used can be varied for additional accuracy or clarity.

The proteins encoded by the extended cDNAs may also be used to generate antibodies as explained in Examples 30 and 43 in order to identify the tissue type or cell species from which a sample is derived as described in example 51.

5

### EXAMPLE 51

#### Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations according to Examples 30 and 43 which are  
10 conjugated, directly or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

Antisera for these procedures must have a potency exceeding that of the native  
15 preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are  
20 labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

#### *A. Immunohistochemical techniques*

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, Chap. 26 in: *Basic and Clinical*  
25 *Immunology*, 3rd Ed. Lange, Los Altos, California, 1980, or Rose, *et al.*, Chap. 12 in: *Methods in Immunodiagnosis*, 2d Ed. John Wiley and Sons, New York (1980), the disclosures of which are incorporated herein by reference.

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such  
30 as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin

or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example  $^{125}\text{I}$ , and detected by overlaying the antibody treated preparation with photographic emulsion.

5       Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required.

10       Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4  $\mu\text{m}$ , unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example,  
15       buffer.

Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed.

20       If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available.

25       The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

#### *B. Identification of tissue specific soluble proteins*

30       The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an

electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection.

A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as  
5 required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis.

A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis, *et al.*,  
10 Section 19-2 in: *Basic Methods in Molecular Biology*, Leder ed., Elsevier, New York, 1986, the disclosure of which is incorporated herein by reference, using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5  
15 to 55  $\mu$ l, and containing from about 1 to 100  $\mu$ g protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis, L. *et al.*, *supra* Section 19-3. One set of nitrocellulose blots is stained with Coomassie blue dye to visualize the entire set of proteins for comparison  
20 with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described in Examples 30 and 43. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

In either procedure A or B, a detectable label can be attached to the primary tissue  
25 antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another  
30 strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody.

The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from extended cDNA sequences, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign  
5 bodily sites.

In addition to their applications in forensics and identification, 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be mapped to their chromosomal locations. Example 52 below describes radiation hybrid (RH) mapping of human chromosomal regions using 5'ESTs. Example 53 below describes a representative  
10 procedure for mapping an 5' EST to its location on a human chromosome. Example 54 below describes mapping of 5' ESTs on metaphase chromosomes by Fluorescence In Situ Hybridization (FISH). Those skilled in the art will appreciate that the method of Examples 52-54 may also be used to map cDNAs or genomic DNAs obtainable from the 5' ESTs to their chromosomal locations.

15

## 2. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof in Chromosome Mapping

### **EXAMPLE 52**

#### Radiation hybrid mapping of 5'ESTs to the human genome

20 Radiation hybrid (RH) mapping is a somatic cell genetic approach that can be used for high resolution mapping of the human genome. In this approach, cell lines containing one or more human chromosomes are lethally irradiated, breaking each chromosome into fragments whose size depends on the radiation dose. These fragments are rescued by fusion with cultured rodent cells, yielding subclones containing different portions of the human  
25 genome. This technique is described by Benham *et al.*, *Genomics* 4:509-517, 1989; and Cox *et al.*, *Science* 250:245-250, 1990, the entire contents of which are hereby incorporated by reference. The random and independent nature of the subclones permits efficient mapping of any human genome marker. Human DNA isolated from a panel of 80-100 cell lines provides a mapping reagent for ordering 5'EST. In this approach, the frequency of breakage between  
30 markers is used to measure distance, allowing construction of fine resolution maps as has

been done using conventional ESTs (Schuler *et al.*, *Science* 274:540-546, 1996, hereby incorporated by reference).

RH mapping has been used to generate a high-resolution whole genome radiation hybrid map of human chromosome 17q22-q25.3 across the genes for growth hormone (GH) and thymidine kinase (TK) (Foster *et al.*, *Genomics* 33:185-192, 1996), the region surrounding the Gorlin syndrome gene (Obermayr *et al.*, *Eur. J. Hum. Genet.* 4:242-245, 1996), 60 loci covering the entire short arm of chromosome 12 (Raeymaekers *et al.*, *Genomics* 29:170-178, 1995), the region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer *et al.*, *Genomics* 14:574-584, 1992) and 13 loci on the long arm of chromosome 5 (Warrington *et al.*, *Genomics* 11:701-708, 1991).

### EXAMPLE 53

#### Mapping of 5'ESTs to Human Chromosomes using PCR techniques

5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich in *PCR Technology, Principles and Applications for DNA Amplification*, Freeman and Co., New York, 1992, the disclosure of which is incorporated herein by reference..

The primers are used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions are as follows: 60 ng of genomic DNA is used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1  $\mu$ Cu of a  $^{32}$ P-labeled deoxycytidine triphosphate. The PCR is performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products are analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the length of the resulting PCR product is identical to the distance between the ends of the primer sequences in the extended cDNA from which the primers are

derived, then the PCR reaction is repeated with DNA templates from two panels of human-rodent somatic cell hybrids, BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

5        PCR is used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given 5' EST (or cDNA or genomic DNA obtainable therefrom). DNA is isolated from the somatic hybrids and used as starting templates for PCR reactions using the primer pairs from the 5' EST (or cDNA or genomic DNA obtainable therefrom). Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the 5' EST (or cDNA or genomic DNA obtainable therefrom) will yield an amplified fragment. The 5' EST (or cDNA or genomic DNA obtainable therefrom) are assigned to a chromosome by analysis of the segregation pattern of PCR products from the somatic hybrid DNA templates. The single human chromosome present in all cell hybrids that give rise to an amplified fragment is the chromosome containing that 5'EST (or cDNA or genomic DNA obtainable therefrom). For a review of techniques and analysis of results from somatic cell gene mapping experiments, see Ledbetter *et al.*, 15        *Genomics* 6:475-481, 1990, the disclosure of which is incorporated herein by reference.

#### EXAMPLE 54

##### Mapping of Extended 5' ESTs to Chromosomes Using Fluorescence *In Situ* 20        Hybridization

Fluorescence in situ hybridization allows the 5'EST (or cDNA or genomic DNA obtainable therefrom) to be mapped to a particular location on a given chromosome. The chromosomes to be used for fluorescence in situ hybridization techniques may be obtained from a variety of sources including cell cultures, tissues, or whole blood.

25        In a preferred embodiment, chromosomal localization of an 5'EST (or cDNA or genomic DNA obtainable therefrom) is obtained by FISH as described by Cherif *et al.* (*Proc. Natl. Acad. Sci. U.S.A.*, 87:6639-6643, 1990), the disclosure of which is incorporated herein by reference. Metaphase chromosomes are prepared from phytohemagglutinin (PHA)-stimulated blood cell donors. PHA-stimulated lymphocytes from healthy males are cultured 30        for 72 h in RPMI-1640 medium. For synchronization, methotrexate (10  $\mu$ M) is added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BrdU, 0.1 mM) for 6 h. Colcemid (1

µg/ml) is added for the last 15 min before harvesting the cells. Cells are collected, washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37°C for 15 min and fixed in three changes of methanol:acetic acid (3:1). The cell suspension is dropped onto a glass slide and air dried. The 5'EST (or cDNA or genomic DNA obtainable therefrom) is labeled with  
5 biotin-16 dUTP by nick translation according to the manufacturer's instructions (Bethesda Research Laboratories, Bethesda, MD), purified using a Sephadex G-50 column (Pharmacia, Upsala, Sweden) and precipitated. Just prior to hybridization, the DNA pellet is dissolved in hybridization buffer (50% formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated salmon sperm DNA, pH 7) and the probe is denatured at 70°C for 5-10 min.

10 Slides kept at -20°C are treated for 1 h at 37°C with RNase A (100 µg/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol series. Chromosome preparations are denatured in 70% formamide, 2 X SSC for 2 min at 70°C, then dehydrated at 4°C. The slides are treated with proteinase K (10 µg/100 ml in 20 mM Tris-HCl, 2 mM CaCl<sub>2</sub>) at 37°C for 8 min and dehydrated. The hybridization mixture containing the probe is placed on the  
15 slide, covered with a coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37°C. After hybridization and post-hybridization washes, the biotinylated probe is detected by avidin-FITC and amplified with additional layers of biotinylated goat anti-avidin and avidin-FITC. For chromosomal localization, fluorescent R-bands are obtained as previously described (Cherif *et al.*, *supra.*). The slides are observed under a LEICA  
20 fluorescence microscope (DMRXA). Chromosomes are counterstained with propidium iodide and the fluorescent signal of the probe appears as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular 5'EST (or cDNA or genomic DNA obtainable therefrom) may be localized to a particular cytogenetic R-band on a given chromosome.

25

Once the 5'EST (or cDNA or genomic DNA obtainable therefrom) have been assigned to particular chromosomes using the techniques described in Examples 52-54 above, they may be utilized to construct a high resolution map of the chromosomes on which they are located or to identify the chromosomes in a sample.

30



**EXAMPLE 55**Use of 5'EST to Construct or Expand Chromosome Maps

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique sequences located on the same chromosome. One approach to chromosome mapping utilizes a series of yeast artificial chromosomes (YACs) bearing several thousand long inserts derived from the chromosomes of the organism from which the extended cDNAs (or genomic DNAs obtainable therefrom) are obtained. This approach is described in Nagaraja *et al.*, *Genome Research* 7:210-222, 1997, the disclosure of which is incorporated herein by reference. Briefly, in this approach each chromosome is broken into overlapping pieces which are inserted into the YAC vector. The YAC inserts are screened using PCR or other methods to determine whether they include the 5'EST (or cDNA or genomic DNA obtainable therefrom) whose position is to be determined. Once an insert has been found which includes the 5'EST (or cDNA or genomic DNA obtainable therefrom), the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences known to be on the chromosome or in the region from which the 5'EST (or cDNA or genomic DNA obtainable therefrom) was derived. This process can be repeated for each insert in the YAC library to determine the location of each of the extended cDNAs (or genomic DNAs obtainable therefrom) relative to one another and to other known chromosomal markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms chromosomes may be obtained.

As described in Example 56 below extended cDNAs (or genomic DNAs obtainable therefrom) may also be used to identify genes associated with a particular phenotype, such as hereditary disease or drug response.

3. Use of 5'ESTs or Sequences Obtained Therefrom or Fragments Thereof in Gene Identification

**EXAMPLE 56****Identification of genes associated with hereditary diseases or drug response**

This example illustrates an approach useful for the association of 5'ESTs (or cDNA or genomic DNA obtainable therefrom) with particular phenotypic characteristics. In this  
5 example, a particular 5'EST (or cDNA or genomic DNA obtainable therefrom) is used as a test probe to associate that 5'EST (or cDNA or genomic DNA obtainable therefrom) with a particular phenotypic characteristic.

5'ESTs (or cDNA or genomic DNA obtainable therefrom) are mapped to a particular location on a human chromosome using techniques such as those described in Examples 52  
10 and 53 or other techniques known in the art. A search of Mendelian Inheritance in Man (McKusick in *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library) reveals the region of the human chromosome which contains the 5'EST (or cDNA or genomic DNA obtainable therefrom) to be a very gene rich region containing several known genes and several diseases or phenotypes for which genes  
15 have not been identified. The gene corresponding to this 5'EST (or cDNA or genomic DNA obtainable therefrom) thus becomes an immediate candidate for each of these genetic diseases.

Cells from patients with these diseases or phenotypes are isolated and expanded in culture. PCR primers from the 5'EST (or cDNA or genomic DNA obtainable  
20 therefrom) are used to screen genomic DNA, mRNA or cDNA obtained from the patients. 5'ESTs (or cDNA or genomic DNA obtainable therefrom) that are not amplified in the patients can be positively associated with a particular disease by further analysis. Alternatively, the PCR analysis may yield fragments of different lengths when the samples are derived from an individual having the phenotype associated with the  
25 disease than when the sample is derived from a healthy individual, indicating that the gene containing the 5'EST may be responsible for the genetic disease.

**VI. Use of 5'EST (or cDNA or Genomic DNA Obtainable Therefrom) to Construct Vectors**

30 The present 5'ESTs (or cDNA or genomic DNA obtainable therefrom) may also be used to construct secretion vectors capable of directing the secretion of the proteins

encoded by genes therein. Such secretion vectors may facilitate the purification or enrichment of the proteins encoded by genes inserted therein by reducing the number of background proteins from which the desired protein must be purified or enriched. Exemplary secretion vectors are described in Example 57 below.

5

#### 1. Construction of Secretion Vectors

##### EXAMPLE 57

##### Construction of Secretion Vectors

The secretion vectors include a promoter capable of directing gene expression in the host cell, tissue, or organism of interest. Such promoters include the Rous Sarcoma Virus promoter, the SV40 promoter, the human cytomegalovirus promoter, and other promoters familiar to those skilled in the art.

A signal sequence from a 5' EST (or cDNAs or genomic DNAs obtainable therefrom) is operably linked to the promoter such that the mRNA transcribed from the promoter will direct the translation of the signal peptide. The host cell, tissue, or organism may be any cell, tissue, or organism which recognizes the signal peptide encoded by the signal sequence in the 5' EST (or cDNA or genomic DNA obtainable therefrom). Suitable hosts include mammalian cells, tissues or organisms, avian cells, tissues, or organisms, insect cells, tissues or organisms, or yeast.

In addition, the secretion vector contains cloning sites for inserting genes encoding the proteins which are to be secreted. The cloning sites facilitate the cloning of the insert gene in frame with the signal sequence such that a fusion protein in which the signal peptide is fused to the protein encoded by the inserted gene is expressed from the mRNA transcribed from the promoter. The signal peptide directs the extracellular secretion of the fusion protein.

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Many nucleic acid backbones suitable for use as secretion vectors are known to those skilled in the art, including retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial chromosomes, P element vectors,

baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

5       After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as  
10   ammonium sulfate precipitation, immunoprecipitation, immunochromatography, size exclusion chromatography, ion exchange chromatography, and HPLC. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

      The signal sequences may also be inserted into vectors designed for gene therapy. In  
15   such vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the promoter encodes the signal peptide. A cloning site is located downstream of the signal sequence such that a gene encoding a protein whose secretion is desired may readily be inserted into the vector and fused to the signal sequence. The vector is introduced into an appropriate host cell. The protein expressed from the promoter is secreted  
20   extracellularly, thereby producing a therapeutic effect.

      The 5' ESTs may also be used to clone sequences located upstream of the 5' ESTs which are capable of regulating gene expression, including promoter sequences, enhancer sequences, and other upstream sequences which influence transcription or  
25   translation levels. Once identified and cloned, these upstream regulatory sequences may be used in expression vectors designed to direct the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative fashion. Example 58 describes a method for cloning sequences upstream of the extended cDNAs or 5' ESTs.

## 2. Identification of Upstream Sequences With Promoting or Regulatory Activities

### **EXAMPLE 58**

#### Use of Extended cDNAs or 5' ESTs to Clone Upstream Sequences from Genomic DNA

Sequences derived from extended cDNAs or 5' ESTs may be used to isolate the  
5 promoters of the corresponding genes using chromosome walking techniques. In one  
chromosome walking technique, which utilizes the GenomeWalker™ kit available from  
Clontech, five complete genomic DNA samples are each digested with a different restriction  
enzyme which has a 6 base recognition site and leaves a blunt end. Following digestion,  
oligonucleotide adapters are ligated to each end of the resulting genomic DNA fragments.

10 For each of the five genomic DNA libraries, a first PCR reaction is performed  
according to the manufacturer's instructions (which are incorporated herein by reference)  
using an outer adaptor primer provided in the kit and an outer gene specific primer. The gene  
specific primer should be selected to be specific for the extended cDNA or 5' EST of interest  
and should have a melting temperature, length, and location in the extended cDNA or 5' EST  
15 which is consistent with its use in PCR reactions. Each first PCR reaction contains 5 ng of  
genomic DNA, 5 µl of 10X Tth reaction buffer, 0.2 mM of each dNTP, 0.2 µM each of outer  
adaptor primer and outer gene specific primer, 1.1 mM of Mg(OAc)<sub>2</sub>, and 1 µl of the Tth  
polymerase 50X mix in a total volume of 50 µl. The reaction cycle for the first PCR reaction  
is as follows: 1 min - 94°C / 2 sec - 94°C, 3 min - 72°C (7 cycles) / 2 sec - 94°C, 3 min -  
20 -67°C (32 cycles) / 5 min - 67°C.

The product of the first PCR reaction is diluted and used as a template for a  
second PCR reaction according to the manufacturer's instructions using a pair of nested  
primers which are located internally on the amplicon resulting from the first PCR  
reaction. For example, 5 µl of the reaction product of the first PCR reaction mixture  
25 may be diluted 180 times. Reactions are made in a 50 µl volume having a composition  
identical to that of the first PCR reaction except the nested primers are used. The first  
nested primer is specific for the adaptor, and is provided with the GenomeWalker™ kit.  
The second nested primer is specific for the particular extended cDNA or 5' EST for  
which the promoter is to be cloned and should have a melting temperature, length, and  
30 location in the extended cDNA or 5' EST which is consistent with its use in PCR  
reactions. The reaction parameters of the second PCR reaction are as follows: 1 min -

94°C / 2 sec - 94°C, 3 min - 72°C (6 cycles) / 2 sec - 94°C, 3 min - 67°C (25 cycles) / 5 min - 67°C. The product of the second PCR reaction is purified, cloned, and sequenced using standard techniques.

Alternatively, two or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors which can be converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising at least 15 nucleotides from the extended cDNA or 5' EST sequence is hybridized to the single stranded DNA. Hybrids between the biotinylated oligonucleotide and the single stranded DNA containing the extended cDNA or EST sequence are isolated as described in Example 29 above. Thereafter, the single stranded DNA containing the extended cDNA or EST sequence is released from the beads and converted into double stranded DNA using a primer specific for the extended cDNA or 5' EST sequence or a primer corresponding to a sequence included in the cloning vector. The resulting double stranded DNA is transformed into bacteria. DNAs containing the 5' EST or extended cDNA sequences are identified by colony PCR or colony hybridization.

Once the upstream genomic sequences have been cloned and sequenced as described above, prospective promoters and transcription start sites within the upstream sequences may be identified by comparing the sequences upstream of the extended cDNAs or 5' ESTs with databases containing known transcription start sites, transcription factor binding sites, or promoter sequences.

In addition, promoters in the upstream sequences may be identified using promoter reporter vectors as described in Example .

25

### EXAMPLE 59

#### Identification of Promoters in Cloned Upstream Sequences

The genomic sequences upstream of the extended cDNAs or 5' ESTs are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline

30

phosphatase,  $\beta$  galactosidase, or green fluorescent protein. The sequences upstream of the extended cDNAs or 5' ESTs are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for augmenting transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

Appropriate host cells for the promoter reporter vectors may be chosen based on the results of the above described determination of expression patterns of the extended cDNAs and ESTs. For example, if the expression pattern analysis indicates that the mRNA corresponding to a particular extended cDNA or 5' EST is expressed in fibroblasts, the promoter reporter vector may be introduced into a human fibroblast cell line.

Promoter sequences within the upstream genomic DNA may be further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

### EXAMPLE 60

#### Cloning and Identification of Promoters

Using the method described in Example 58 above with 5' ESTs, sequences upstream of several genes were obtained. Using the primer pairs GGG AAG ATG GAG ATA GTA TTG CCT G (SEQ ID NO:29) and CTG CCA TGT ACA TGA TAG AGA GAT TC (SEQ

ID NO:30), the promoter having the internal designation P13H2 (SEQ ID NO:31) was obtained.

Using the primer pairs GTA CCA GGGG ACT GTG ACC ATT GC (SEQ ID NO:32) and CTG TGA CCA TTG CTC CCA AGA GAG (SEQ ID NO:33), the promoter  
5 having the internal designation P15B4 (SEQ ID NO:34) was obtained.

Using the primer pairs CTG GGA TGG AAG GCA CGG TA (SEQ ID NO:35) and GAG ACC ACA CAG CTA GAC AA (SEQ ID NO:36), the promoter having the internal designation P29B6 (SEQ ID NO:37) was obtained.

Figure 4 provides a schematic description of the promoters isolated and the way they  
10 are assembled with the corresponding 5' tags. The upstream sequences were screened for the presence of motifs resembling transcription factor binding sites or known transcription start sites using the computer program MatInspector release 2.0, August 1996.

Table VII describes the transcription factor binding sites present in each of these promoters. The columns labeled matrice provides the name of the MatInspector matrix used.  
15 The column labeled position provides the 5' position of the promoter site. Numeration of the sequence starts from the transcription site as determined by matching the genomic sequence with the 5' EST sequence. The column labeled "orientation" indicates the DNA strand on which the site is found, with the + strand being the coding strand as determined by matching the genomic sequence with the sequence of the 5' EST. The column labeled "score" provides  
20 the MatInspector score found for this site. The column labeled "length" provides the length of the site in nucleotides. The column labeled "sequence" provides the sequence of the site found.

Bacterial clones containing plasmids containing the promoter sequences described above described above are presently stored in the inventor's laboratories under the internal  
25 identification numbers provided above. The inserts may be recovered from the deposited materials by growing an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a  
30 cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography.

The plasmid DNA obtained using these procedures may then be manipulated using standard



cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the EST insertion. The PCR product which corresponds to the 5' EST can then be manipulated using standard cloning techniques familiar to those skilled in the art.

5       The promoters and other regulatory sequences located upstream of the extended cDNAs or 5' ESTs may be used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative manner. A promoter capable of directing the desired spatial, temporal, developmental, and quantitative patterns may be selected using the results of the expression analysis described in  
10   Example 26 above. For example, if a promoter which confers a high level of expression in muscle is desired, the promoter sequence upstream of an extended cDNA or 5' EST derived from an mRNA which is expressed at a high level in muscle, as determined by the method of Example 26, may be used in the expression vector.

      Preferably, the desired promoter is placed near multiple restriction sites to facilitate  
15   the cloning of the desired insert downstream of the promoter, such that the promoter is able to drive expression of the inserted gene. The promoter may be inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine  
20   Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

      Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

      Following the identification of promoter sequences using the procedures of Examples  
25   58-60, proteins which interact with the promoter may be identified as described in Example 61 below.

**EXAMPLE 61****Identification of Proteins Which Interact with Promoter Sequences, Upstream  
Regulatory Sequences, or mRNA**

Sequences within the promoter region which are likely to bind transcription factors  
5 may be identified by homology to known transcription factor binding sites or through  
conventional mutagenesis or deletion analyses of reporter plasmids containing the promoter  
sequence. For example, deletions may be made in a reporter plasmid containing the promoter  
sequence of interest operably linked to an assayable reporter gene. The reporter plasmids  
carrying various deletions within the promoter region are transfected into an appropriate host  
10 cell and the effects of the deletions on expression levels is assessed. Transcription factor  
binding sites within the regions in which deletions reduce expression levels may be further  
localized using site directed mutagenesis, linker scanning analysis, or other techniques familiar  
to those skilled in the art.

Nucleic acids encoding proteins which interact with sequences in the promoter  
15 may be identified using one-hybrid systems such as those described in the manual  
accompanying the Matchmaker One-Hybrid System kit available from Clontech (Catalog  
No. K1603-1), the disclosure of which is incorporated herein by reference. Briefly, the  
Matchmaker One-hybrid system is used as follows. The target sequence for which it is  
desired to identify binding proteins is cloned upstream of a selectable reporter gene and  
20 integrated into the yeast genome. Preferably, multiple copies of the target sequences are  
inserted into the reporter plasmid in tandem. A library comprised of fusions between  
cDNAs to be evaluated for the ability to bind to the promoter and the activation domain  
of a yeast transcription factor, such as GAL4, is transformed into the yeast strain  
containing the integrated reporter sequence. The yeast are plated on selective media to  
25 select cells expressing the selectable marker linked to the promoter sequence. The  
colonies which grow on the selective media contain genes encoding proteins which bind  
the target sequence. The inserts in the genes encoding the fusion proteins are further  
characterized by sequencing. In addition, the inserts may be inserted into expression  
vectors or *in vitro* transcription vectors. Binding of the polypeptides encoded by the  
30 inserts to the promoter DNA may be confirmed by techniques familiar to those skilled in  
the art, such as gel shift analysis or DNase protection analysis.

## VII. Use of 5' ESTs (or cDNAs or Genomic DNAs Obtainable Therefrom) in Gene Therapy

The present invention also comprises the use of 5'ESTs (or cDNA or genomic DNA obtainable therefrom) in gene therapy strategies, including antisense and triple helix strategies as described in Examples 62 and 63 below. In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense sequences may prevent gene expression through a variety of mechanisms. For example, the antisense sequences may inhibit the ability of ribosomes to translate the mRNA. Alternatively, the antisense sequences may block transport of the mRNA from the nucleus to the cytoplasm, thereby limiting the amount of mRNA available for translation. Another mechanism through which antisense sequences may inhibit gene expression is by interfering with mRNA splicing. In yet another strategy, the antisense nucleic acid may be incorporated in a ribozyme capable of specifically cleaving the target mRNA.

### EXAMPLE 62

#### Preparation and Use of Antisense Oligonucleotides

The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. They may comprise a sequence complementary to the sequence of the 5'EST (or cDNA or genomic DNA obtainable therefrom). The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex with sufficient stability to inhibit the expression of the mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green *et al.*, *Ann. Rev. Biochem.* 55:569-597, 1986; and Izant and Weintraub, *Cell* 36:1007-1015, 1984, which are hereby incorporated by reference.

In some strategies, antisense molecules are obtained from a nucleotide sequence encoding a protein by reversing the orientation of the coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using *in vitro* transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach

involves transcription of the antisense nucleic acids *in vivo* by operably linking DNA containing the antisense sequence to a promoter in an expression vector.

Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be synthesized *in vitro*. Thus, the antisense nucleic acids are complementary to the corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use in antisense strategies are described by Rossi *et al.*, *Pharmacol. Ther.* **50(2)**:245-254, 1991, which is hereby incorporated by reference.

Various types of antisense oligonucleotides complementary to the sequence of the 5'EST (or cDNA or genomic DNA obtainable therefrom) may be used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides described in International Application No. PCT WO94/23026, hereby incorporated by reference, are used. In these molecules, the 3' end or both the 3' and 5' ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense oligonucleotides.

In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141, hereby incorporated by reference, are used.

In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523, hereby incorporated by reference, are used. These double- or single-stranded oligonucleotides comprise one or more, respectively, inter- or intra-oligonucleotide covalent cross-linkages, wherein the linkage consists of an amide bond between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand, respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide monosaccharide ring, and the carboxyl group being carried by an aliphatic spacer group substituted on a nucleotide or nucleotide analog of the other strand or the same strand, respectively.

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO 92/18522, incorporated by reference, may also be used. These molecules are stable to degradation and contain at least one transcription control recognition sequence which binds to control proteins and are effective as decoys therefore. These molecules may contain "hairpin" structures, "dumbbell" structures, "modified dumbbell" structures, "cross-linked" decoy structures and "loop" structures.

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in European Patent Application No. 0 572 287 A2, hereby incorporated by reference are used. These ligated oligonucleotide "dumbbells" contain the binding site for a transcription factor and inhibit expression of the gene under control of the transcription factor by sequestering the factor.

Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732, hereby incorporated by reference, is also contemplated. Because these molecules have no free ends, they are more resistant to degradation by exonucleases than are conventional oligonucleotides. These oligonucleotides may be multifunctional, interacting with several regions which are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit gene expression may be determined using *in vitro* expression analysis. The antisense molecule may be introduced into the cells by diffusion, injection, infection, transfection or h-region-mediated import using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector may be any of a variety of expression vectors known in the art, including retroviral or viral vectors, vectors capable of extrachromosomal replication, or integrating vectors. The vectors may be DNA or RNA.

The antisense molecules are introduced onto cell samples at a number of different concentrations preferably between  $1 \times 10^{-10}$  M to  $1 \times 10^{-4}$  M. Once the minimum concentration that can adequately control gene expression is identified, the optimized dose is translated into a dosage suitable for use *in vivo*. For example, an inhibiting concentration in culture of  $1 \times 10^{-7}$  translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide

approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate.

5 It is further contemplated that the antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to specifically bind and cleave its target mRNA. For technical applications of ribozyme and antisense oligonucleotides see Rossi *et al.*, *supra*.

10 In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the effectiveness of antisense inhibition on translation can be monitored using techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling.

The 5' ESTs of the present invention (or cDNAs or genomic DNAs obtainable therefrom) may also be used in gene therapy approaches based on intracellular triple helix formation. Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity as it is associated with a particular gene. The 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene expression in individuals having diseases associated with expression of a particular gene. Similarly, a portion of 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) can be used to study the effect of inhibiting transcription of a particular gene within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the 5'EST or from the gene corresponding to the 5'EST are contemplated within the scope of this invention.

15  
20  
25

**EXAMPLE 63**Preparation and Use of Triple Helix Probes

The sequences of the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) are scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting gene expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting gene expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which normally express the target gene. The oligonucleotides may be prepared on an oligonucleotide synthesizer or they may be purchased commercially from a company specializing in custom oligonucleotide synthesis, such as GENSET, Paris, France.

The oligonucleotides may be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced gene expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the target gene in cells which have been treated with the oligonucleotide. The cell functions to be monitored are predicted based upon the homologies of the target gene corresponding to the extended cDNA from which the oligonucleotide was derived with known gene sequences that have been associated with a particular function. The cell functions can also be predicted based on the presence of abnormal physiologies within cells derived from individuals with a particular inherited disease, particularly when the extended cDNA is associated with the disease using techniques described in Example 56.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced *in vivo* using the techniques described above and in Example 62 at a dosage calculated based on the *in vitro* results, as described in Example 62.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the

generation of oligonucleotides suitable for triple helix formation see Griffin *et al.*, *Science* 245:967-971, 1989, which is hereby incorporated by this reference.

### EXAMPLE 64

5 Use of cDNAs Obtained Using the 5' ESTs to Express an Encoded Protein in a Host Organism

The cDNAs obtained as described above using the 5' ESTs of the present invention may also be used to express an encoded protein in a host organism to produce a beneficial effect. In such procedures, the encoded protein may be transiently expressed in the host organism or stably expressed in the host organism. The encoded protein may have any of the activities described above. The encoded protein may be a protein which the host organism lacks or, alternatively, the encoded protein may augment the existing levels of the protein in the host organism.

15 A full length extended cDNA encoding the signal peptide and the mature protein, or an extended cDNA encoding only the mature protein is introduced into the host organism. The extended cDNA may be introduced into the host organism using a variety of techniques known to those of skill in the art. For example, the extended cDNA may be injected into the host organism as naked DNA such that the encoded protein is expressed in the host organism, thereby producing a beneficial effect.

Alternatively, the extended cDNA may be cloned into an expression vector downstream of a promoter which is active in the host organism. The expression vector may be any of the expression vectors designed for use in gene therapy, including viral or retroviral vectors. The expression vector may be directly introduced into the host organism such that the encoded protein is expressed in the host organism to produce a beneficial effect. In another approach, the expression vector may be introduced into cells *in vitro*. Cells containing the expression vector are thereafter selected and introduced into the host organism, where they express the encoded protein to produce a beneficial effect.



**EXAMPLE 65**

Use of Signal Peptides Encoded by 5' ESTs or Sequences obtained Therefrom  
to Import Proteins Into Cells

The short core hydrophobic region (h) of signal peptides encoded by the 5'ESTS or  
5 extended cDNAs derived from SEQ ID NOs: 38-185 may also be used as a carrier to import  
a peptide or a protein of interest, so-called cargo, into tissue culture cells (Lin *et al.*, *J. Biol.  
Chem.*, **270**: 14225-14258, 1995; Du *et al.*, *J. Peptide Res.*, **51**: 235-243, 1998; Rojas *et al.*,  
*Nature Biotech.*, **16**: 370-375, 1998).

When cell permeable peptides of limited size (approximately up to 25 amino acids)  
10 are to be translocated across cell membrane, chemical synthesis may be used in order to add  
the h region to either the C-terminus or the N-terminus to the cargo peptide of interest.  
Alternatively, when longer peptides or proteins are to be imported into cells, nucleic acids can  
be genetically engineered, using techniques familiar to those skilled in the art, in order to link  
the extended cDNA sequence encoding the h region to the 5' or the 3' end of a DNA  
15 sequence coding for a cargo polypeptide. Such genetically engineered nucleic acids are then  
translated either *in vitro* or *in vivo* after transfection into appropriate cells, using conventional  
techniques to produce the resulting cell permeable polypeptide. Suitable hosts cells are then  
simply incubated with the cell permeable polypeptide which is then translocated across the  
membrane.

20 This method may be applied to study diverse intracellular functions and cellular  
processes. For instance, it has been used to probe functionally relevant domains of  
intracellular proteins and to examine protein-protein interactions involved in signal  
transduction pathways (Lin *et al.*, *supra*; Lin *et al.*, *J. Biol. Chem.*, **271**: 5305-5308, 1996;  
Rojas *et al.*, *J. Biol. Chem.*, **271**: 27456-27461, 1996; Liu *et al.*, *Proc. Natl. Acad. Sci. USA*,  
25 **93**: 11819-11824, 1996; Rojas *et al.*, *Bioch. Biophys. Res. Commun.*, **234**: 675-680, 1997).

Such techniques may be used in cellular therapy to import proteins producing  
therapeutic effects. For instance, cells isolated from a patient may be treated with imported  
therapeutic proteins and then re-introduced into the host organism.

Alternatively, the h region of signal peptides of the present invention could be used in  
30 combination with a nuclear localization signal to deliver nucleic acids into cell nucleus. Such  
oligonucleotides may be antisense oligonucleotides or oligonucleotides designed to form

triple helixes, as described in examples 62 and 63 respectively, in order to inhibit processing and/or maturation of a target cellular RNA.

As discussed above, the cDNAs or portions thereof obtained using the 5' ESTs of the present invention can be used for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris *et al.*, *Cell* 75:791-803, 1993, the disclosure of which is hereby incorporated by reference) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins

involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

5           Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation *Molecular Cloning; A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, Fritsch and Maniatis eds., 1989, and *Methods in Enzymology; Guide to Molecular Cloning Techniques*, Academic Press, Berger and Kimmel eds., 1987.

10           Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid  
15 preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the  
20 disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims. All documents cited herein are incorporated herein by reference in their entirety.

Step	Search characteristic		Selection Characteristics		
	Program	Strand	Parameters	Identity (%)	Length (bp)
miscellaneous	blastn	both	S=61 X=16	90	17
tRNA	fasta	both	-	80	60
rRNA	blastn	both	S=108	80	40
mtRNA	blastn	both	S=108	80	40
Prokaryotic	blastn	both	S=144	90	40
Fungal	blastn	both	S=144	90	40
Alu	fasta*	both	-	70	40
L1	blastn	both	S=72	70	40
Repeats	blastn	both	S=72	70	40
Promoters	blastn	top	S=54 X=16	90	15†
Vertebrate	fasta*	both	S=108	90	30
ESTs	blastn	both	S=108 X=16	90	30
Proteins	blastx <sup>□</sup>	top	E = 0.001	-	-

Table 1: Parameters used for each step of EST analysis

\* use "Quick Fast" Database scanner

† alignment further constrained to begin closer than 10bp to EST's end

□ using BLOSUM62 substitution matrix

TABLE II

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID38	new	11.8	Umbilical cord	37-4-1-A12-PU
ID39	new	10	Lymph ganglia	48-50-1-G11-PU
ID40	new	10	Lymph ganglia	48-16-2-C11-PU
ID41	new	10	Placenta	14-8-1-C10-PU
ID42	new	9.9	Lymph ganglia	48-48-3-E11-PU
ID43	new	9.6	Lymph ganglia	48-26-2-B9-PU
ID44	new	9.2	Lymph ganglia	48-25-4-D9-PU
ID45	new	8.9	Lymph ganglia	48-67-2-F5-PU
ID46	new	8.9	Lymph ganglia	48-47-4-H7-PU
ID47	new	8.6	Lymph ganglia	48-52-1-E10-PU
ID48	new	8.5	Placenta	14-8-4-G8-PU
ID49	new	8.4	Lymph ganglia	48-4-2-G5-PU
ID50	new	8.2	Lymph ganglia	48-27-2-D7-PU
ID51	new	7	Lymph ganglia	48-61-3-F5-PU
ID52	new	6.9	Placenta	14-7-4-G8
ID53	new	6.9	Lymph ganglia	48-5-4-B6-PU
ID54	new	6.8	Lymph ganglia	48-46-3-C8-PU
ID55	new	6.7	Lymph ganglia	48-20-3-A6-PU
ID56	new	6.6	Lymph ganglia	48-18-2-F6-PU
ID57	new	6.5	Lymph ganglia	48-2-2-A10-PU
ID58	new	6.5	Lymph ganglia	48-25-4-C11-PU
ID59	new	6.3	Lymph ganglia	48-26-1-C4-PU
ID60	new	6.3	Lymph ganglia	48-31-2-G8-PU
ID61	new	6.3	Lymph ganglia	48-24-1-D8-PU
ID62	new	6.3	Umbilical cord	37-11-2-D10-PU
ID63	new	6.3	Lymph ganglia	48-8-2-C2-PU
ID64	new	6.2	Lymph ganglia	48-20-4-A8-PU
ID65	new	6.1	Lymph ganglia	48-2-1-B9-PU
ID66	new	6.1	Lymph ganglia	48-54-1-G9-PU
ID67	new	6.1	Lymph ganglia	48-47-4-B7-PU
ID68	new	6.1	Lymph ganglia	48-8-1-D8-PU
ID69	new	5.9	Lymph ganglia	48-12-3-G8-PU
ID70	new	5.9	Umbilical cord	37-39-4-A5-PU
ID71	new	5.9	Lymph ganglia	48-25-1-B6-PU
ID72	new	5.7	Lymph ganglia	48-15-1-D2-PU
ID73	new	5.5	Umbilical cord	37-3-4-D1-PU
ID74	new	5.4	Lymph ganglia	48-13-1-G4-PU
ID75	new	5.4	Lymph ganglia	48-10-1-E4-PU
ID76	new	5.2	Lymph ganglia	48-8-1-A3-PU
ID77	new	5.2	Umbilical cord	37-7-4-F2-PU
ID78	new	5.2	Lymph ganglia	48-50-3-F1-PU
ID79	new	5.2	Lymph ganglia	48-8-2-B5-PU
ID80	new	5	Placenta	11-4-0-B11-RP
ID81	new	5	Lymph ganglia	48-48-4-H11-PU
ID82	new	4.9	Umbilical cord	37-2-1-B4-PU
ID83	new	4.8	Lymph ganglia	48-47-2-B2-PU
ID84	new	4.8	Lymph ganglia	48-3-4-C11-PU
ID85	new	4.8	Lymphocytes	24-6-1-C8-PU
ID86	new	4.8	Placenta	31-10-3-D2-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID87	new	4.7	Lymph ganglia	48-54-3-F9-PU
ID88	new	4.7	Lymph ganglia	48-47-1-C9-PU
ID89	new	4.7	Lymph ganglia	48-4-2-C9-PU
ID90	new	4.6	Umbilical cord	37-33-2-E2-PU
ID91	new	4.5	Umbilical cord	37-2-1-B7-PU
ID92	new	4.5	Lymph ganglia	48-51-2-C3-PU
ID93	new	4.5	Lymph ganglia	48-23-4-D4-PU
ID94	new	4.4	Umbilical cord	37-4-1-A6-PU
ID95	new	4.4	Lymph ganglia	48-11-4-C10-PU
ID96	new	4.4	Umbilical cord	37-1-4-F3-PU
ID97	new	4.3	Lymphocytes	24-2-2-G10-PU
ID98	new	4.3	Lymph ganglia	48-26-3-G3-PU
ID99	new	4.1	Lymph ganglia	48-20-3-H2-PU
ID100	new	4.1	Lymph ganglia	48-31-3-F7-PU
ID101	new	4.1	Lymph ganglia	48-29-1-H9-PU
ID102	new	4.1	Umbilical cord	37-1-3-G4-PU
ID103	new	4.1	Umbilical cord	37-8-3-G12-PU
ID104	new	4.1	Lymph ganglia	48-26-4-G1-PU
ID105	new	4	Lymph ganglia	48-27-1-B12-PU
ID106	new	4	Lymph ganglia	48-22-1-H7-PU
ID107	new	4	Lymphocytes	24-1-4-F9-PU
ID108	new	4	Lymph ganglia	48-6-2-A1-PU
ID109	new	4	Umbilical cord	37-3-3-B3-PU
ID110	new	3.8	Umbilical cord	37-7-2-F6-PU
ID111	new	3.8	Lymph ganglia	48-52-1-A6-PU
ID112	new	3.8	Lymph ganglia	48-7-2-F5-PU
ID113	new	3.8	Umbilical cord	37-12-2-D12-PU
ID114	new	3.8	Umbilical cord	37-11-3-D2-PU
ID115	new	3.8	Lymph ganglia	48-1-1-H7-PU
ID116	new	3.7	Lymph ganglia	48-21-3-E1-PU
ID117	new	3.6	Lymph ganglia	48-26-3-B8-PU
ID118	new	3.6	Umbilical cord	37-9-2-D9-PU
ID119	new	3.6	Lymph ganglia	48-3-3-A3-PU
ID120	new	3.6	Lymphocytes	24-1-3-G11-PU
ID121	new	3.6	Lymphocytes	24-4-1-A4-PU
ID122	new	3.5	Lymph ganglia	48-23-2-B12-PU
ID123	new	3.5	Lymph ganglia	48-47-3-F2-PU
ID124	new	3.5	Lymphocytes	24-4-4-H11-PU
ID125	new	3.5	Lymph ganglia	48-7-3-B8-PU
ID126	ext-est-not-vrt	12.8	Lymph ganglia	48-12-4-E3-PU
ID127	ext-est-not-vrt	9.3	Umbilical cord	37-12-3-G9-PU
ID128	ext-est-not-vrt	9.3	Lymph ganglia	48-67-4-A6-PU
ID129	ext-est-not-vrt	8.1	Lymph ganglia	48-28-3-A9-PU
ID130	ext-est-not-vrt	7.7	Lymphocytes	24-3-3-C6-PU
ID131	ext-est-not-vrt	6.6	Lymph ganglia	48-28-4-C2-PU
ID132	ext-est-not-vrt	6.2	Lymph ganglia	48-25-2-A1-PU
ID133	ext-est-not-vrt	5.8	Lymph ganglia	48-24-4-B7-PU
ID134	ext-est-not-vrt	5.3	Lymph ganglia	48-6-1-C9-PU
ID135	ext-est-not-vrt	5.1	Lymph ganglia	48-7-4-H2-PU
ID136	ext-est-not-vrt	4.6	Lymph ganglia	48-28-3-B6-PU
ID137	ext-est-not-vrt	4.4	Lymph ganglia	48-3-1-H9-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID138	ext-est-not-vrt	4.4	Umbilical cord	37-6-4-B11-PU
ID139	ext-est-not-vrt	3.9	Lymph ganglia	48-26-1-G10-PU
ID140	ext-est-not-vrt	3.8	Umbilical cord	37-9-4-H9-PU
ID141	ext-est-not-vrt	3.7	Lymphocytes	24-1-4-F8-PU
ID142	ext-est-not-vrt	3.5	Lymph ganglia	48-21-3-H7-PU
ID143	est-not-ext	11.7	Lymph ganglia	48-6-4-G3-PU
ID144	est-not-ext	10.9	Umbilical cord	37-5-1-A12-PU
ID145	est-not-ext	10.9	Lymph ganglia	48-22-4-A8-PU
ID146	est-not-ext	9.6	Lymph ganglia	48-27-1-B8-PU
ID147	est-not-ext	9.6	Umbilical cord	37-4-1-G3-PU
ID148	est-not-ext	9.3	Lymph ganglia	48-11-4-E3-PU
ID149	est-not-ext	8.2	Lymph ganglia	48-25-4-D8-PU
ID150	est-not-ext	8.2	Lymph ganglia	48-19-3-G1-PU
ID151	est-not-ext	8.1	Placenta	31-11-4-B2-PU
ID152	est-not-ext	7.9	Lymph ganglia	48-7-4-H10-PU
ID153	est-not-ext	7.7	Lymph ganglia	48-11-4-F7-PU
ID154	est-not-ext	7.2	Lymph ganglia	48-10-3-B5-PU
ID155	est-not-ext	6.9	Umbilical cord	37-8-4-D3-PU
ID156	est-not-ext	6.4	Umbilical cord	37-6-2-D10-PU
ID157	est-not-ext	6.3	Lymph ganglia	48-17-1-D11-PU
ID158	est-not-ext	6.1	Lymphocytes	24-8-3-G1-PU
ID159	est-not-ext	6.1	Umbilical cord	37-12-2-D1-PU
ID160	est-not-ext	6.1	Umbilical cord	37-6-2-A10-PU
ID161	est-not-ext	6	Lymph ganglia	48-26-1-A11-PU
ID162	est-not-ext	5.9	Lymph ganglia	48-60-4-H5-PU
ID163	est-not-ext	5.8	Umbilical cord	37-29-2-G3-PU
ID164	est-not-ext	5.7	Umbilical cord	37-28-2-D3-PU
ID165	est-not-ext	5.6	Lymph ganglia	48-49-1-F5-PU
ID166	est-not-ext	5.5	Umbilical cord	37-2-2-D12-PU
ID167	est-not-ext	5.5	Umbilical cord	37-7-4-B3-PU
ID168	est-not-ext	5.3	Lymph ganglia	48-24-1-D2-PU
ID169	est-not-ext	5	Lymph ganglia	48-21-4-H4-PU
ID170	est-not-ext	4.9	Umbilical cord	37-41-4-B9-PU
ID171	est-not-ext	4.9	Lymph ganglia	48-12-3-E2-PU
ID172	est-not-ext	4.6	Lymph ganglia	48-5-4-C5-PU
ID173	est-not-ext	4.3	Lymphocytes	24-5-1-E2-PU
ID174	est-not-ext	4.1	Lymph ganglia	48-18-3-F9-PU
ID175	est-not-ext	4.1	Lymphocytes	24-5-1-H2-PU
ID176	est-not-ext	3.8	Lymph ganglia	48-6-2-G1-PU
ID177	est-not-ext	3.8	Umbilical cord	37-9-2-G10-PU
ID178	est-not-ext	3.7	Lymph ganglia	48-19-3-A7-PU
ID179	est-not-ext	3.5	Lymph ganglia	48-13-3-E3-PU
ID180	est-not-ext	3.5	Lymph ganglia	48-20-4-G6-PU
ID181	est-not-ext	3.5	Lymphocytes	24-4-1-G11-PU
ID182	est-not-ext	3.5	Lymph ganglia	48-4-2-E4-PU
ID183	ext-vrt-not-genomic	8.4	Lymph ganglia	48-24-1-B3-PU
ID184	ext-vrt-not-genomic	7.4	Lymph ganglia	48-30-2-B2-PU
ID185	ext-vrt-not-genomic	6.5	Umbilical cord	37-30-2-B3-PU

TABLE III

<u>SEQ. ID</u> <u>NO.</u>	<u>SIGNAL PEPTIDE</u>
ID38	MVLVALILLHSALA
ID39	MAQHHLWILLCLQTWPEAAG
ID40	MKDLWIFLLVTAPRCILS
ID41	MAQHHLWILLCLQTWPEAAG
ID42	MDWTWRFLFVVAATGVQS
ID43	MSICFLGLLLCLPHRLA
ID44	MIGFLVLLILPLLSSLS
ID45	MQCLLSVLMAQFIXHFLSLLMSLLVSTVTWQ
ID46	MELGLSWIFFLATLKGVCQ
ID47	MVSVSLALLSGWVGS
ID48	MPLPWSLALPLLLSWVAGGFG
ID49	MVSNFFHVIQVFEKSATLISKTEHIGFVITYSWXKSTTHLGSRKFAISYILSEVSLQKYD CPFSGTSFVVFSFLICAMA
ID50	MRXFWFLMYPFRFHDCKQKYDLYISIAGWLIICLACVLFPLLRT
ID51	MVSLCCLFTCFIPICIS
ID52	MDFFFLERSYWGKMILLVTYSPIAYS
ID53	MTMRHNWTPDLSPLWVLLCAHVVTL
ID54	MDNMSGGKVDEALVKSSCLHPWSKRNDVSMQCSQDILRMLLSLQPVLO
ID55	MXLQQEATGKVLKIHKDTSQVPTAXGDASIAALVWTLPGAQR
ID56	MTEHSLTHQGIPILVLILFPTSCVM
ID57	MYIGGLRFIFLTSQLISS
ID58	MSVSLKHHHLHFIMSVLVFWNCSHLIFFSLIFLNLFA
ID59	MXXLGXRFMVVSFLSXPFLCSA
ID60	MDWTWYILVSVAATGAHS
ID61	MISKFSSKAYSVRGLELFSLLPINSPNSAIXVACVLSSLI AVNS
ID62	MVLLGAFGSCIKSFSLLFLIFSLNLNRG
ID63	MAARQAVGSGAQETCGLDRILEALKLLSPGXSGS
ID64	MSTQKGLALFLMALGFSCI
ID65	MKDVEIIMIFHGYFLIVFFVFLCNC
ID66	MCFPEHRRQMYIQDRLDVTRRARQGRICALLLQSQCAYWA
ID67	MLVVKQCFSDSSILSTFVSWLSA
ID68	MIXLRDTAASRLERDTRQLPLLTSAHGLQQ
ID69	MITMMLALISVCLF
ID70	MWLLTLVQCSDLCP
ID71	MRVHLFPYLCQPSVLSNLLFACLTMLLVKT
ID72	MIPLCFLILPYPVLS
ID73	MAGSRLPRQLFLQGVXASSCLLXPSTRKSQA
ID74	MYICFCLESFEIKCGFVLHLLAQDLVCC
ID75	MHFILHNLNAFTLLVWLSLS
ID76	MSFFPFNRSLNSNPHNLLFPNIAPLFTLLPKSIP
ID77	MVVWVLEVRFLDLHCFCSLAKT
ID78	MVCGWWTQGPVPGLCPPALGSAWS
ID79	MGRAFPSRHKTARFECALVSASLTTA
ID80	MGLKALCXXLLCVLFVSH
ID81	MMATQTLSDSYQDQGMQVVTTELKTEQDPNCSEPD AEGVSPPPVESQTPMDVDKQAIYR HPLFPLLALLFEKCEQ
ID82	MSPSQLTCSVFLSGSVCLSFL
ID83	MLQALAPAHHLCSLKRSFCSLLCLRTQLFP
ID84	MLFLKYLWRS�CRG



<u>SEQ. ID</u> <u>NO.</u>	<u>SIGNAL PEPTIDE</u>
ID85	MALLAMHSWRWAAAAAAFEKRRHSAILRPLVSVSGS
ID86	MKAXAMFGAGDEDDTDFLSPSGGARLASLFGLDQXAAG
ID87	MLWLLRSLTDVSS
ID88	MTIFHVLIHSSSFS
ID89	MHWQLLXGFCGSYSA
ID90	MTMMVMASFLPRNTMYTNTMNYSIFVLLFFFSXLXY
ID91	MPSQTLSQPRISVLHGDLVPAGMAVQEIGAQMVLPCVVSGSGLTREHLVTRLALCQS PRA
ID92	MSLRVHTLPTLLGAVVRPGCRELLCLLMITVTVGPGAS
ID93	MYLTSLLLLGRWLTLS
ID94	MNWNVRGTRGFLLCPLVCGLRR
ID95	MEQAALVVSPLPRRCSVRSPVTTCCAKDLVCLTFITATTHE
ID96	MIPLPSLVGCWEGGNGKGLMVSDTTCTLASSNVSPSPAPTGLGRGAPSHTPQKKPTIP GARHRPIILPKGLVQLHATXLALG
ID97	MSMRLSGERIYLLLEVWLPXLNFESVLHFIQTVHIALPGSLG
ID98	MGTLFFFMPVVPV
ID99	MVVLPMTLGIYLLQFLSIVS
ID100	MAPHTASFGVCPLLSVTRVVATEHWLFLASLSGIKT
ID101	MSYKWMPSLPCLSFCTCLV
ID102	MPLPTWAPTLAGFLLVLYVCLP
ID103	MNLYLLDWIGLKALIRG
ID104	MSCXVXDAXXRWWAHXLIIGWXHLTQKVHPIALSHCVNMGTLLFFCFMPVVPV
ID105	MVPNLCGRQILAFQTFLLNLRA
ID106	MFSLIIFFPSSP
ID107	MSAFYLSYLLHCLLIVFILVEF
ID108	MAEAKLVQGSVAPQRXSAGVVTMDGASA
ID109	MKGVGPEQLNDGAPSNEIEMTPCFSEFLLLDVGVVNIVVIKMSYNNVLLTISTNASVLG
ID110	MLRKLSASNENCLLSNPISHNEVYLIRCCESHQLFWVTASTFCRS
ID111	MYPLILLPLNPFVLQ
ID112	MLLRSPGSPRGFVAVGLGQISA
ID113	MARPGATACGPAAHQCSA
ID114	MEPVSSSLCIXXLEHLFT
ID115	MRPAGRWCASAAWRSPLSA
ID116	MWLCAYVLFFFNGCLY
ID117	MLLLHRAVVLRLQQA
ID118	MEMFGXXEKDFSSVEGVXLXSLVPSMCFHVTNS
ID119	MQMHGWRWDPHSSEQDLAHTLSREASLENNTALLGVHASFQMSVA
ID120	MASPRGTDYNQTPNTTMYCYAVGTGVLTSLRLARA
ID121	MAPILSSFKSLLKYHLETSLSILLKPVTLLHCLCPFPALFLS
ID122	MNRLSKHLILVPWWLPFVYT
ID123	MSSNKEQRSAVFVILFALITILLYSSNS
ID124	MDMKSNTHGHLFLGRQPSFSVRSMGPALAIQCQPHNPGPPMGTPTEDPSGCSFPCLFLS PQSFLVLS
ID125	MSEAGCKPSRPEHGSFLSLSTLLLTSHH
ID126	MESGXGXVFLVALLRGVQC
ID127	MLCRLFTLLLLQSLLLG
ID128	MDLLHKNMKHLWFFLLL VAGPRWVLS
ID129	MQAQAPVVVVTPQGVGPGPAPQNSNWQTMCDGSDCGVCLCGTFCFPCLG
ID130	MKALCLLLLPVLGLLVSS
ID131	MSPSGRLCLLXIVGLXLPXG
ID132	MLLAWVQAFVSNMMLAEAYG

SEQ. ID NO.	SIGNAL PEPTIDE
ID133	MLSESRGPPVQEHEAPVVLPPAGGGSQMGPVPAAXAGESGPGXVKPLETLXLTCVSGGS IS
ID134	MTSGQARASXQSPQALEDSGPVNISVSITLTLDPLKPFGGYSRNVTHLYSTILGHQIGLS GREAAHEEINITFTLPTAWSSDDCALHGHCEQVVFTACMTLTASPGVFP
ID135	MLGGDHRALLLKIWLLQRPES
ID136	MRFRKAWAPVLAALSHSLMSLLDESSCQA
ID137	MYVWPCAVVLAQYLWFHRRSLPGKAILEIGAGVSLPGILAAKCGAEVILSDSSELPHCLE VCRQSCQMNNLPHLQVVGLTWG
ID138	MLNPAQXDTMPCEYLSLDAMEKWIIFGILCHGILNTXATALNLWKLALQSSSCLS
ID139	MNAQASSSRCHGVCLSVPSLPSIS
ID140	MAKVQVNNVVLDNPPFYNPFQFEITFECIEDLSEDLWKITYVGSAAESEEYDQVLDSV LVGPVPA
ID141	MADVEDGEETCALASHSGSSG
ID142	MWTCLLGDCGPPEA
ID143	MDWTWXVFCLLA VAPGAHS
ID144	MDNSWRLGPAIGLSAGQSQLLVSLLLLLTRVQP
ID145	MXHLXFFLLLVAAPRWVLS
ID146	MPVPASWPHPPGPFLLL TLLGLTEVAG
ID147	MKEYVLLLFLALCSA
ID148	MAQSLALSLLLVLAFG
ID149	MKKVLLLITAILAVAVG
ID150	MKKVLLLITAILAVAVG
ID151	MRIMLLFTAILAFSLAQS
ID152	MAWTVLLLGLLSHCTVS
ID153	MTILHTGXNPFRRPSQRWTAPALLHHRPXTXPPSXHRSRCTEXVGIPXLLLQTL PASTX
ID154	MKHLWFFLLLVAAPKXXLS
ID155	MLSYFLSSLVCGSLGLSNVSG
ID156	MGTQDPQAEQGLRIPLPGLLLSKHHHPAPELPALALLHAGHA
ID157	MMTIYALSNEFAFKINEEQLSXXPLXSVQLXHA
ID158	MARGAHLXALEMLTAFASHIRA
ID159	MNPESPQQLERQSTGPRTGTRRCLSKFTWCTSRMMTQTCILLIHTMQVCTT
ID160	MMTQTCILLIHTMQVCTT
ID161	MAGKGSSGRPLLLGLLVAVATVHL
ID162	MAGSPTCLTLIYLWQLTGSA
ID163	MVGMVCFILGLIICQC
ID164	MXLLHSLSSGVRA
ID165	MTMAECPTLCVSSSPALWA
ID166	MVPLVAVVSGPRAQLFACLLRLGTQ
ID167	MSEMAELSELYEESDLQMDVMPGEGDLPQMEVGSGSRELSLRPSRSGAQQL EE EGP MEE EEAQPMAXQRGNGALLTGPTLGSSQA
ID168	MLIVSVLALIPXTT
ID169	MTCRGSCSYATTRSPSELSLLPSSLWVLA
ID170	MEAVVFVFSLLDCCA
ID171	MAATSGTDEPVSGELVSVAHALSLPAQSYG
ID172	MADEALFLLHNMVSG
ID173	MASMQKRLQKELLALQNDPPPGMTLNEKSVQNSITQWIVDMEGAPGTLYEGEKFQLLFKF SSRYPDFSPQVMFTGENIPVHPHVYSNGHICLSILTEDWSPALSVQSVCLSIISMLSSC
ID174	MKXMTGSENWKTCKVLMFCVTPPELET
ID175	MQHIVGVPHVLVRRGLLGRDLFMTRTLCSPGPGS
ID176	MYHQSEALALASSQSHLLG
ID177	MSGQGLAGFFASVAMICAIASG

SEQ. ID  
NO.

SIGNAL PEPTIDE

ID178	MPTGKQLADIGYKTFSTSMMLLTVYGGYLC
ID179	MFPVCLTVTAAVCG
ID180	MSVIFACVVRVRDG
ID181	MLXGGLKMAPRGKRLSSTPLEILFFLNGWYNATYFLELFIFLYKGVLLPYPTANLVLDV VMLLLYLG
ID182	MIGGGRWDPPGAQAPSSQAFRRPALTIHLPGTEG
ID183	MVRRVQPDRKQLPLVLLRLLCLLPTGLP
ID184	MPLHYSLVFIIGLVGNLLA
ID185	MARGLGAPHWVAVGLLTWATLGLLVAGLGG

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3.5	0.121	0.036	0.467	0.664
4	0.096	0.06	0.519	0.708
4.5	0.078	0.079	0.565	0.745
5	0.062	0.098	0.615	0.782
5.5	0.05	0.127	0.659	0.813
6	0.04	0.163	0.694	0.836
6.5	0.033	0.202	0.725	0.855
7	0.025	0.248	0.763	0.878
7.5	0.021	0.304	0.78	0.889
8	0.015	0.368	0.816	0.909
8.5	0.012	0.418	0.836	0.92
9	0.009	0.512	0.856	0.93
9.5	0.007	0.581	0.863	0.934
10	0.006	0.679	0.835	0.919

TABLE IV

125/2

Minimum signal peptide score	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
3.5	2674	947	599	23	150
4	2278	784	499	23	126
4.5	1943	647	425	22	112
5	1657	523	353	21	96
5.5	1417	419	307	19	80
6	1190	340	238	18	68
6.5	1035	280	186	18	60
7	893	219	161	15	48
7.5	753	173	132	12	36
8	636	133	101	11	29
8.5	543	104	83	8	26
9	456	81	63	6	24
9.5	364	57	48	6	18
10	303	47	35	6	15

TABLE V

Tissue	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
Brain	329	131	75	3	24
Cancerous prostate	134	40	37	1	6
Cerebellum	17	9	1	0	6
Colon	21	11	4	0	0
Dystrophic muscle	41	18	8	0	1
Fetal brain	70	37	16	0	1
Fetal kidney	227	116	46	1	19
Fetal liver	13	7	2	0	0
Heart	30	15	7	0	1
Hypertrophic prostate	86	23	22	2	2
Kidney	10	7	3	0	0
Large intestine	21	8	4	0	1
Liver	23	9	6	0	0
Lung	24	12	4	0	1
Lung (cells)	57	38	6	0	4
Lymph ganglia	163	60	23	2	12
Lymphocytes	23	6	4	0	2
Muscle	33	16	6	0	4
Normal prostate	181	61	45	7	11
Ovary	90	57	12	1	2
Pancreas	48	11	6	0	1
Placenta	24	5	1	0	0
Prostate	34	16	4	0	2
Spleen	56	28	10	0	1
Substantia nigra	108	47	27	1	6
Surrenals	15	3	3	1	0
Testis	131	68	25	1	8
Thyroid	17	8	2	0	2
Umbilical cord	55	17	12	1	3
Uterus	28	15	3	0	2
Non tissue-specific	568	48	177	2	28
Total	2677	947	601	23	150

TABLE VI

125/4

# Description of Transcription Factor Binding Sites present on promoters isolated from SignalTag sequences

Promoter sequence P13H2 (646 bp):

Matrix	Position	Orientation	Score	Length	Sequence
CMYB_01	-502	+	0.983	9	TGTCAGTTG
MYOD_Q6	-501	-	0.961	10	CCCAACTGAC
S8_01	-444	-	0.960	11	AATAGAATTAG
S8_01	-425	+	0.966	11	AACTAAATTAG
DELTAEF1_01	-390	-	0.980	11	GCACACCTCAG
GATA_C	-364	-	0.964	11	AGATAAATCCA
CMYB_01	-349	+	0.958	9	CTTCAGTTG
GATA1_02	-343	+	0.959	14	TTGTAGATAGGACA
GATA_C	-339	+	0.953	11	AGATAGGACAT
TAL1ALPHA47_01	-235	+	0.973	16	CATAACAGATGGTAAG
TAL1BETA47_01	-235	+	0.983	16	CATAACAGATGGTAAG
TAL1BETA1F2_01	-235	+	0.978	16	CATAACAGATGGTAAG
MYOD_Q6	-232	-	0.954	10	ACCATCTGTT
GATA1_04	-217	-	0.953	13	TCAAGATAAAGTA
IK1_01	-126	+	0.963	13	AGTTGGGAATTCC
IK2_01	-126	+	0.985	12	AGTTGGGAATTC
CREL_01	-123	+	0.962	10	TGGGAATTCC
GATA1_02	-96	+	0.950	14	TCAGTGATATGGCA
SRY_02	-41	-	0.951	12	TAAAACAAAACA
E2F_02	-33	+	0.957	8	TTTAGCGC
MZF1_01	-5	-	0.975	8	TGAGGGGA

Promoter sequence P15B4 (861bp) :

Matrix	Position	Orientation	Score	Length	Sequence
NFY_Q6	-748	-	0.956	11	GGACCAATCAT
MZF1_01	-738	+	0.962	8	CCTGGGGA
CMYB_01	-684	+	0.994	9	TGACCGTTG
VMYB_02	-682	-	0.985	9	TCCAACGGT
STAT_01	-673	+	0.968	9	TTCTTGGA
STAT_01	-673	-	0.951	9	TTCCAGGA
MZF1_01	-556	-	0.958	8	TTGGGGGA
IK2_01	-451	+	0.965	12	GAATGGGATTTCC
MZF1_01	-424	+	0.986	8	AGAGGGGA
SRY_02	-398	-	0.955	12	GAAAACAAAACA
MZF1_01	-216	+	0.960	8	GAAGGGGA
MYOD_Q6	-190	+	0.981	10	AGCATCTGCC
DELTAEF1_01	-176	+	0.958	11	TCCCACCTTCC
S8_01	5	-	0.992	11	GAGGCAATTAT
MZF1_01	16	-	0.986	8	AGAGGGGA

Promoter sequence P29B6 (555 bp) :

Matrix	Position	Orientation	Score	Length	Sequence
ARNT_01	-311	+	0.964	16	GGACTCACGTGCTGCT
NMYC_01	-309	+	0.965	12	ACTCACGTGCTG
USF_01	-309	+	0.985	12	ACTCACGTGCTG
USF_01	-309	-	0.985	12	CAGCACGTGAGT
NMYC_01	-309	-	0.956	12	CAGCACGTGAGT
MYCMAX_02	-309	-	0.972	12	CAGCACGTGAGT
USF_C	-307	+	0.997	8	TCACGTGC
USF_C	-307	-	0.991	8	GCACGTGA
MZF1_01	-292	-	0.968	8	CATGGGGA
ELK1_02	-105	+	0.963	14	CTCTCCGGAAGCCT
CETS1P54_01	-102	+	0.974	10	TCCGGAAGCC
AP1_Q4	-42	-	0.963	11	AGTGACTGAAC
AP1FJ_Q2	-42	-	0.961	11	AGTGACTGAAC
PADS_C	45	+	1.000	9	TGTGGTCTC

TABLE VII

CLAIMS

1. A purified or isolated nucleic acid comprising the sequence of one of SEQ ID NOs: 38-185 or comprising a sequence complementary thereto.
- 5 2. The nucleic acid of Claim 1, wherein said nucleic acid is recombinant.
3. A purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-185 or one of the sequences complementary thereto.
4. A purified or isolated nucleic acid comprising at least 15 consecutive bases of  
10 one of the sequences of SEQ ID NOs: 38-185 or one of the sequences complementary thereto.
5. The nucleic acid of Claim 4, wherein said nucleic acid is recombinant.
6. A purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-185 or one of the  
15 sequences complementary to the sequences of SEQ ID NOs: 38-185.
7. The nucleic acid of Claim 6, wherein said nucleic acid is recombinant.
8. A purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-185.
- 20 9. A purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-185 or having a sequence complementary thereto.
10. A purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 38-185 which encode a signal peptide.
11. A purified or isolated polypeptides comprising a signal peptide encoded by  
25 one of the sequences of SEQ ID NOs: 38-185.
12. A vector encoding a fusion protein comprising a polypeptide and a signal peptide, said vector comprising a first nucleic acid encoding a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-185 operably linked to a second nucleic acid encoding a polypeptide.
- 30 13. A method of directing the extracellular secretion of a polypeptide or the insertion of a polypeptide into the membrane comprising the steps of:



obtaining a vector according to Claim 12; and  
introducing said vector into a host cell such that said fusion protein is secreted into the extracellular environment of said host cell or inserted into the membrane of said host cell.

14. A method of importing a polypeptide into a cell comprising contacting said  
5 cell with a fusion protein comprising a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-185 operably linked to said polypeptide.

15. A method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-185, comprising the steps of:

obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-185;  
10 contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-185 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA;

identifying a cDNA which hybridizes to said detectable probe; and

isolating said cDNA which hybridizes to said probe.

16. An isolated or purified cDNA encoding a human secretory protein, said  
15 human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-185 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 15.

17. The cDNA of Claim 16 wherein said cDNA comprises the full protein coding  
20 sequence partially included in one of the sequences of SEQ ID NOs: 38-185.

18. A method of making a cDNA comprising one of the sequences of SEQ ID NOs: 38-185, comprising the steps of:

contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA;

25 hybridizing said first primer to said polyA tail;

reverse transcribing said mRNA to make a first cDNA strand;

making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-185; and

30 isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

19. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-185 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 18.

5 20. The cDNA of Claim 19 wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-185.

21. The method of Claim 18, wherein the second cDNA strand is made by:  
contacting said first cDNA strand with a first pair of primers, said first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the  
10 sequences of SEQ ID NOs 38-185 and a third primer having a sequence therein which is included within the sequence of said first primer;

performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product;

contacting said first PCR product with a second pair of primers, said second pair of  
15 primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NO:s 38-185 , and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and

performing a second polymerase chain reaction, thereby generating a second PCR product.

20 22. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-185, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 21.

23. The cDNA of Claim 22 wherein said cDNA comprises the full protein coding  
25 sequence partially included in one of the sequences of SEQ ID NOs: 38-185.

24. The method of Claim 18 wherein the second cDNA strand is made by:  
contacting said first cDNA strand with a second primer comprising at least 15  
consecutive nucleotides of the sequences of SEQ ID NOs: 38-185;

hybridizing said second primer to said first strand cDNA; and

30 extending said hybridized second primer to generate said second cDNA strand.

25. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-185 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 24.

5 26. The cDNA of Claim 25, wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-185.

27. A method of making a protein comprising one of the sequences of SEQ ID NO: 186-333, comprising the steps of:

obtaining a cDNA encoding the full protein sequence partially included in one of the  
10 sequences of sequence of SEQ ID NO: 38-185;

inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter;

introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and

15 isolating said protein.

28. An isolated protein obtainable by the method of Claim 27.

29. A method of obtaining a promoter DNA comprising the steps of:

obtaining DNAs located upstream of the nucleic acids of SEQ ID NO: 38-185 or the sequences complementary thereto;

20 screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and

isolating said DNA comprising said identified promoter.

30. The method of Claim 29, wherein said obtaining step comprises chromosome walking from said nucleic acids of SEQ ID NO: 38-185 or sequences complementary thereto.

25 31. The method of Claim 30, wherein said screening step comprises inserting said upstream sequences into a promoter reporter vector.

32. The method of Claim 30, wherein said screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

30 33. An isolated promoter obtainable by the method of Claim 32.

34. An isolated or purified protein comprising one of the sequences of SEQ ID NO: 186-333.

35. In an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length, the improvement comprising inclusion in said array of at least one of the sequences of  
5 SEQ ID NOs: 38-185, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-185, or a fragment thereof of at least 15 consecutive nucleotides.

36. The array of Claim 35 including therein at least two of the sequences of SEQ ID NOs: 38-185, the sequences complementary to the sequences of SEQ ID NOs: 38-185, or fragments thereof of at least 15 consecutive nucleotides.

10 37. The array of Claim 35 including therein at least five of the sequences of SEQ ID NOs: 38-185, the sequences complementary to the sequences of SEQ ID NOs: 38-185, or fragments thereof of at least 15 consecutive nucleotides.

1/4

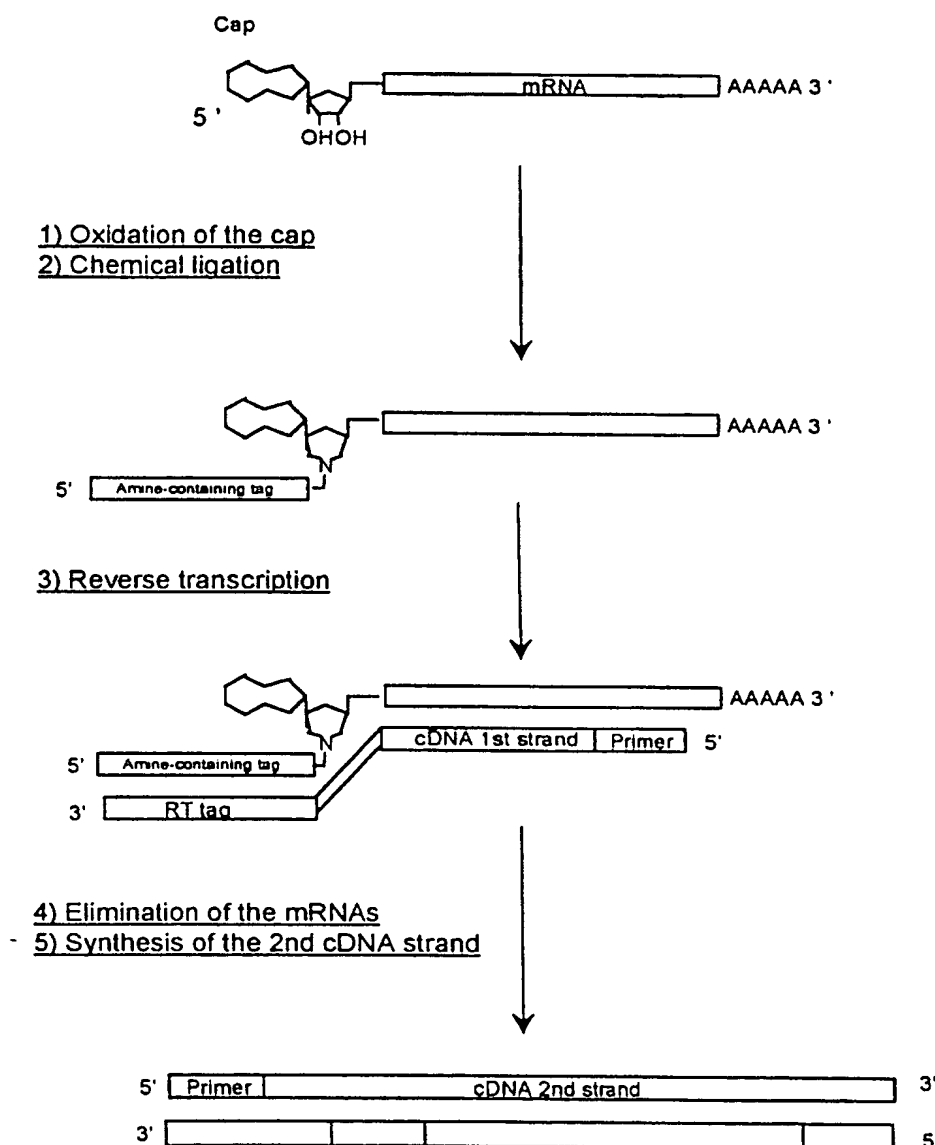


Figure 1

2/4

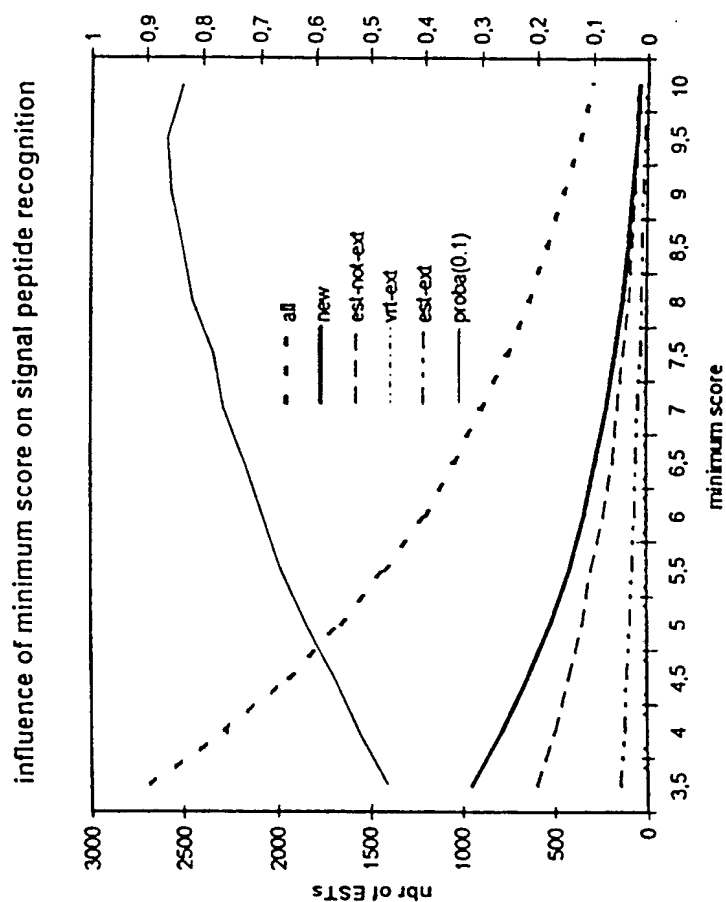


Figure 2

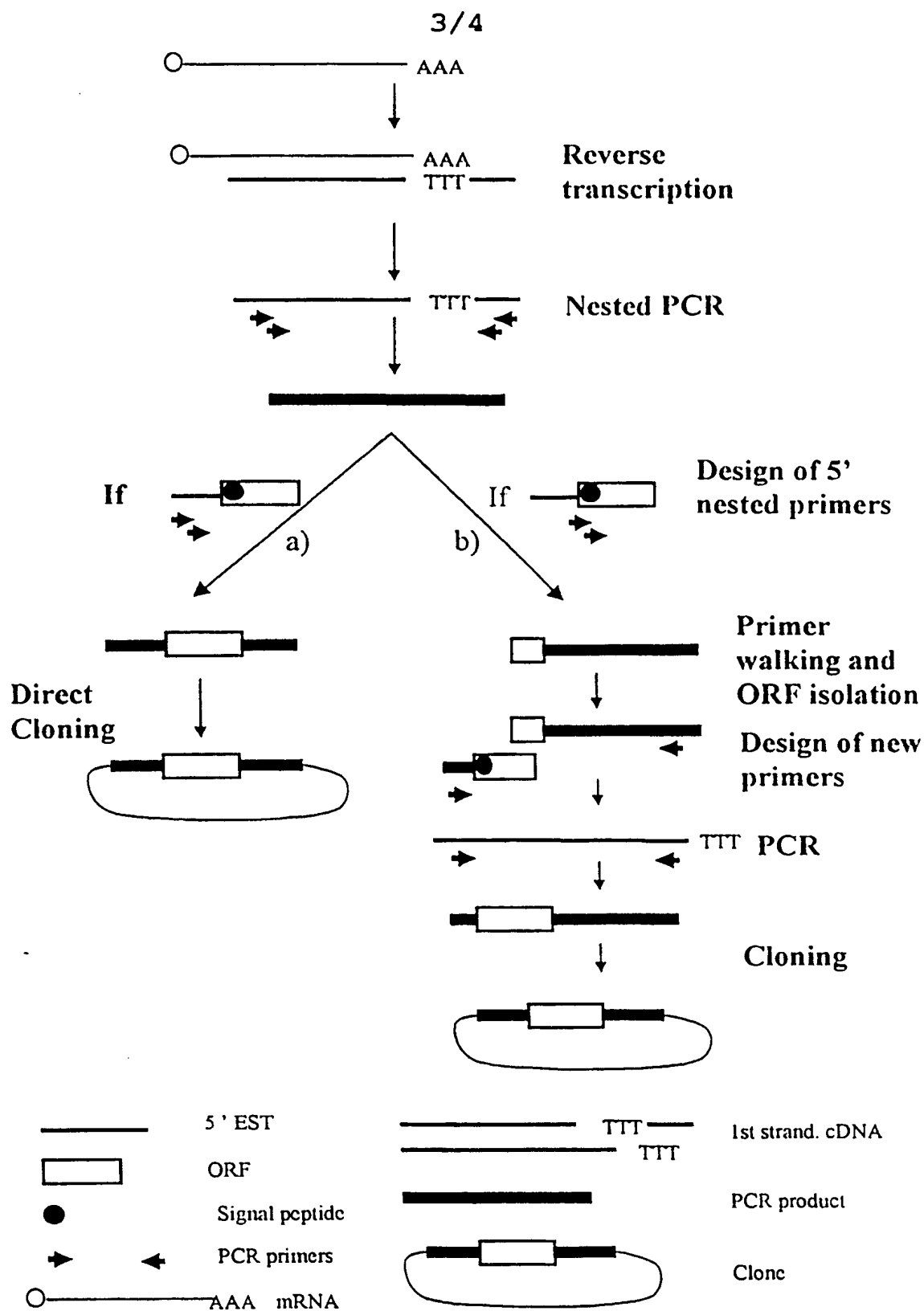
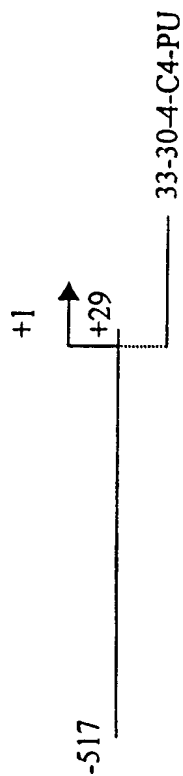
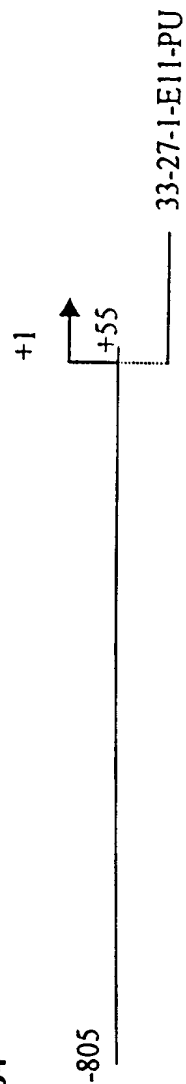


Figure 3

## Promoter P13H2



## Promoter P15B4



## Promoter P29B6

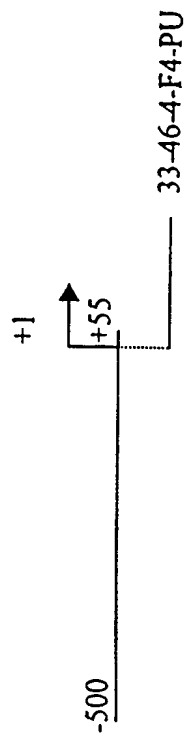


Figure 4



## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (A) NAME: GENSET SA
- (B) STREET: 24, RUE ROYALE
- (C) CITY: PARIS
- (E) COUNTRY: FRANCE
- (F) POSTAL CODE (ZIP) : 75008

(ii) TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS  
EXPRESSED IN VARIOUS TISSUES

(iii) NUMBER OF SEQUENCES: 333

## (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy Disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: Win95
- (D) SOFTWARE: Word

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

## (ix) FEATURE:

- (A) NAME/KEY: Cap
- (B) LOCATION: 1
- (D) OTHER INFORMATION: m7Gppp added to 1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGCAUCCUAC UCCCAUCCAA UCCACCCUA ACUCCUCCCA UCUCAC

47

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GCAUCCUACU CCCAUCCAAU UCCACCCUAA CUCCUCCCAU CUCCAC

46

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATCAAGAATT CGCACGAGAC CATTA

25

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TAATGGTCTC GTGCGAATTC TTGAT

25

## (2) INFORMATION FOR SEQ ID NO: 5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CCGACAAGAC CAACGTCAAG GCCGC

25

## (2) INFORMATION FOR SEQ ID NO: 6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TCACCAGCAG GCAGTGGCTT AGGAG

25

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

AGTGATTCTT GCTACTTTGG ATGGC

25

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCTTGGTCTT GTTCTGGAGT TTAGA

25

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TCCAGAATGG GAGACAAGCC AATTT

25

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AGGGAGGAGG AACAGCGTG AGTCC

25

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: NUCLEIC ACID
  - (C) STRANDEDNESS: SINGLE
  - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATGGGAAAGG AAAAGACTCA TATCA

25

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: NUCLEIC ACID
  - (C) STRANDEDNESS: SINGLE
  - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AGCAGCAACA ATCAGGACAG CACAG

25

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: NUCLEIC ACID
  - (C) STRANDEDNESS: SINGLE
  - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATCAAGAATT CGCACGAGAC CATT

25

## (2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: Other nucleic acid

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ATCGTTGAGA CTCGTACCAG CAGAGTCACG AGAGAGACTA CACGGTACTG GTTTTTTTTTT 60

TTTTTVN 67

## (2) INFORMATION FOR SEQ ID NO: 15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: Other nucleic acid

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CCAGCAGAGT CACGAGAGAG ACTACACGG 29

## (2) INFORMATION FOR SEQ ID NO: 16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: Other nucleic acid

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CACGAGAGAG ACTACACGGT ACTGG 25

## (2) INFORMATION FOR SEQ ID NO: 17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 526 base pairs

- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(261..376)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 166..281  
id N70479  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(380..486)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 54..160  
id N70479  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(110..145)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 403..438  
id N70479  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(196..229)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 315..348  
id N70479  
est

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 90..140
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.2  
seq LLLITAILAVAVG/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```
AATATRARAC AGCTACAATA TTCCAGGGCC ARTCACTTGC CATTTCTCAT AACAGCGTCA      60
GAGAGAAAGA ACTGACTGAR ACGTTTGAG ATG AAG AAA GTT CTC CTC CTG ATC      113
Met Lys Lys Val Leu Leu Leu Ile
-15                               -10
```

ACA GCC ATC TTG GCA GTG GCT GTW GGT TTC CCA GTC TCT CAA GAC CAG	161
Thr Ala Ile Leu Ala Val Ala Val Gly Phe Pro Val Ser Gln Asp Gln	
-5 1 5	
GAA CGA GAA AAA AGA AGT ATC AGT GAC AGC GAT GAA TTA GCT TCA GGR	209
Glu Arg Glu Lys Arg Ser Ile Ser Asp Ser Asp Glu Leu Ala Ser Gly	
10 15 20	
WTT TTT GTG TTC CCT TAC CCA TAT CCA TTT CGC CCA CTT CCA CCA ATT	257
Xaa Phe Val Phe Pro Tyr Pro Tyr Pro Phe Arg Pro Leu Pro Pro Ile	
25 30 35	
CCA TTT CCA AGA TTT CCA TGG TTT AGA CGT AAN TTT CCT ATT CCA ATA	305
Pro Phe Pro Arg Phe Pro Trp Phe Arg Arg Xaa Phe Pro Ile Pro Ile	
40 45 50 55	
CCT GAA TCT GCC CCT ACA ACT CCC CTT CCT AGC GAA AAG TAAACAARAA	354
Pro Glu Ser Ala Pro Thr Thr Pro Leu Pro Ser Glu Lys	
60 65	
GGAAAAGTCA CRATAAACCT GGTCACCTGA AATTGAAATT GAGCCACTTC CTTGAARAAT	414
CAAAATTCCT GTTAATAAAA RAAAAACAAA TGTAATTGAA ATAGCACACA GCATTCTCTA	474
GTCAATATCT TTAGTGATCT TCTTTAATAA ACATGAAAGC AAAAAAAAAA AA	526

## (2) INFORMATION FOR SEQ ID NO: 18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 1..17
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.2  
seq LLLITAILAVAVG/FP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met	Lys	Lys	Val	Leu	Leu	Leu	Ile	Thr	Ala	Ile	Leu	Ala	Val	Ala	Val
1				5				10						15	
Gly															

## (2) INFORMATION FOR SEQ ID NO: 19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 822 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 260..464
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 153..357  
id H57434  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 118..184
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 98..164  
id H57434  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 56..113
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 35..92  
id H57434  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 454..485
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 348..379  
id H57434  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 118..545
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..428  
id N27248  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 65..369
- (C) IDENTIFICATION METHOD: blastn



(D) OTHER INFORMATION: identity 98  
region 41..345  
id H94779  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 61..399  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 6..344  
id H09880  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 408..458  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 92  
region 355..405  
id H09880  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 60..399  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 56..395  
id H29351  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 393..432  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 391..430  
id H29351  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 346..408  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.5  
seq SFLPSALVIWTS/AF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ACTCCTTTTA GCATAGGGGC TTCGGCGCCA GCGGCCAGCG CTAGTCGGTC TGGTAAGTGC	60
CTGATGCCGA GTTCCGTCTC TCGCGTCTTT TCCTGGTCCC AGGCAAAGCG GASGNAGATC	120
CTCAAACGGC CTAGTGCTTC GCGCTCCGG AGAAAATCAG CGGTCTAATT AATTCCTCTG	180
GTTTGTTGAA GCAGTTACCA AGAATCTTCA ACCCTTTCCC ACAAAGCTA ATTGAGTACA	240
CGTTCCTGTT GAGTACACGT TCCTGTTGAT TTACAAAAGG TGCAGGTATG AGCAGGTCTG	300

AAGACTAACA TTTTGTGAAG TTGTAAAACA GAAAACCTGT TAGAA ATG TGG TGG TTT 357  
Met Trp Trp Phe  
-20

CAG CAA GGC CTC AGT TTC CTT CCT TCA GCC CTT GTA ATT TGG ACA TCT 405  
Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val Ile Trp Thr Ser  
-15 -10 -5

GCT GCT TTC ATA TTT TCA TAC ATT ACT GCA GTA ACA CTC CAC CAT ATA 453  
Ala Ala Phe Ile Phe Ser Tyr Ile Thr Ala Val Thr Leu His His Ile  
1 5 10 15

GAC CCG GCT TTA CCT TAT ATC AGT GAC ACT GGT ACA GTA GCT CCA RAA 501  
Asp Pro Ala Leu Pro Tyr Ile Ser Asp Thr Gly Thr Val Ala Pro Xaa  
20 25 30

AAA TGC TTA TTT GGG GCA ATG CTA AAT ATT GCG GCA GTT TTA TGT CAA 549  
Lys Cys Leu Phe Gly Ala Met Leu Asn Ile Ala Ala Val Leu Cys Gln  
35 40 45

AAA TAGAAATCAG GAARATAATT CAACTTAAAG AAKTTCATTT CATGACCAAA 602  
Lys

CTCTTCARAA ACATGTCTTT ACAAGCATAT CTCTTGATT GCTTTCTACA CTGTTGAATT 662

GTCTGGCAAT ATTTCTGCAG TGGAAAATTT GATTARMTA GTTCTTGACT GATAAATATG 722

GTAAGGTGGG CTTTTCCCCC TGTGTAATTG GCTACTATGT CTTACTGAGC CAAGTTGTAW 782

TTTGAAATAA AATGATATGA GAGTGACACA AAAAAAAAAA 822

## (2) INFORMATION FOR SEQ ID NO: 20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 1..21
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.5  
seq SFLPSALVIWTS/AF

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Trp Trp Phe Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val  
1 5 10 15

Ile Trp Thr Ser Ala  
20

## (2) INFORMATION FOR SEQ ID NO: 21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 405 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Testis

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(103..398)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 1..296  
id AA442893  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 185..295
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9  
seq LSYASSALSPCLT/AP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

```

ATCACCTTCT TCTCCATCCT TSTCTGGGCC AGTCCCCARC CCAGTCCCTC TCCTGACCTG      60
CCCAGCCCAA GTCAGCCTTC AGCAGCGCT TTTCTGCACA CAGATATTCC AGGCCTACCT      120
GGCATTCCAG GACCTCCGMA ATGATGCTCC AGTCCCTTAC AAGCGCTTCC TGGATGAGGG      180
TGGC ATG GTG CTG ACC ACC CTC CCC TTG CCC TCT GCC AAC AGC CCT GTG      229
    Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val
        -35                -30                -25

AAC ATG CCC ACC ACT GGC CCC AAC AGC CTG AGT TAT GCT AGC TCT GCC      277
Asn Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala
    -20                -15                -10

CTG TCC CCC TGT CTG ACC GCT CCA AAK TCC CCC CGG CTT GCT ATG ATG      325
Leu Ser Pro Cys Leu Thr Ala Pro Xaa Ser Pro Arg Leu Ala Met Met
    -5                1                5                10

CCT GAC AAC TAAATATCCT TATCCAAATC AATAAARWRA RAATCCTCCC TCCARAAGGG      384
Pro Asp Asn

TTTCTAAAAA CAAAAAAAAA A      405

```

## (2) INFORMATION FOR SEQ ID NO: 22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 1..37
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9  
seq LSYASSALSPCLT/AP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val Asn  
1 5 10 15

Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala Leu  
20 25 30

Ser Pro Cys Leu Thr  
35

## (2) INFORMATION FOR SEQ ID NO: 23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 496 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 149..331
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..183  
id AA397994  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 328..485
- (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 96  
region 179..336  
id AA397994  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(182..496)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 14..328  
id AA399680  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 196..240  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.5  
seq ILSTVTALTFXA/LD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

```

AAAAAATTGG TCCCAGTTTT CACCCTGCCG CAGGGCTGGC TGGGGAGGGC AGCGGTTTAG      60
ATTAGCCGTG GCCTAGGCCG TTTAACGGGG TGACACGAGC NTGCAGGGCC GAGTCCAAGG      120
CCCGGAGATA GGACCAACCG TCAGGAATGC GAGGAATGTT TTTCTTCGGA CTCTATCGAG      180
GCACACAGAC AGACC ATG GGG ATT CTG TCT ACA GTG ACA GCC TTA ACA TTT      231
          Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe
          -15                -10                -5

GCC ARA GCC CTG GAC GGC TGC AGA AAT GGC ATT GCC CAC CCT GCA AGT      279
Ala Xaa Ala Leu Asp Gly Cys Arg Asn Gly Ile Ala His Pro Ala Ser
          1                5                10

GAG AAG CAC AGA CTC GAG AAA TGT AGG GAA CTC GAG ASC ASC CAC TCG      327
Glu Lys His Arg Leu Glu Lys Cys Arg Glu Leu Glu Xaa Xaa His Ser
          15                20                25

GCC CCA GGA TCA ACC CAS CAC CGA AGA AAA ACA ACC AGA AGA AAT TAT      375
Ala Pro Gly Ser Thr Xaa His Arg Arg Lys Thr Thr Arg Arg Asn Tyr
          30                35                40                45

TCT TCA GCC TGAAATGAAK CCGGGATCAA ATGGTTGCTG ATCARAGCCC ATATTTAAAT      434
Ser Ser Ala

TGGAAAAGTC AAATTGASCA TTATTAAATA AAGCTTGTTT AATATGTCTC AAACAAAAAA      494
AA                                                                 496

```

## (2) INFORMATION FOR SEQ ID NO: 24:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids  
(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: 1..15

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 5.5  
seq ILSTVTALTFAXA/LD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe Ala Xaa Ala  
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 623 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Testis

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: 49..96

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 10.1  
seq LVLTLCTLPLAVA/SA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AAAGATCCCT GCAGCCCGGC AGGAGAGAAG GCTGAGCCTT CTGGCGTC ATG GAG AGG 57  
Met Glu Arg  
-15

CTC GTC CTA ACC CTG TGC ACC CTC CCG CTG GCT GTG GCG TCT GCT GGC 105  
Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala Ser Ala Gly  
-10 -5 1

TGC GCC ACG ACG CCA GCT CGC AAC CTG AGC TGC TAC CAG TGC TTC AAG 153  
Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln Cys Phe Lys  
5 10 15

GTC AGC AGC TGG ACG GAG TGC CCG CCC ACC TGG TGC AGC CCG CTG GAC 201  
Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser Pro Leu Asp  
20 25 30 35

CAA GTC TGC ATC TCC AAC GAG GTG GTC GTC TCT TTT AAA TGG AGT GTA	249
Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Lys Trp Ser Val	
40 45 50	
CGC GTC CTG CTC AGC AAA CGC TGT GCT CCC AGA TGT CCC AAC GAC AAC	297
Arg Val Leu Leu Ser Lys Arg Cys Ala Pro Arg Cys Pro Asn Asp Asn	
55 60 65	
ATG AAK TTC GAA TGG TCG CCG GCC CCC ATG GTG CAA GGC GTG ATC ACC	345
Met Xaa Phe Glu Trp Ser Pro Ala Pro Met Val Gln Gly Val Ile Thr	
70 75 80	
AGG CGC TGC TGT TCC TGG GCT CTC TGC AAC AGG GCA CTG ACC CCA CAG	393
Arg Arg Cys Cys Ser Trp Ala Leu Cys Asn Arg Ala Leu Thr Pro Gln	
85 90 95	
GAG GGG CGC TGG GCC CTG CRA GGG GGG CTC CTG CTC CAG GAC CCT TCG	441
Glu Gly Arg Trp Ala Leu Xaa Gly Gly Leu Leu Leu Gln Asp Pro Ser	
100 105 110 115	
AGG GGC ARA AAA ACC TGG GTG CGG CCA CAG CTG GGG CTC CCA CTC TGC	489
Arg Gly Xaa Lys Thr Trp Val Arg Pro Gln Leu Gly Leu Pro Leu Cys	
120 125 130	
CTT CCC AWT TCC AAC CCC CTC TGC CCA RGG GAA ACC CAG GAA GGA	534
Leu Pro Xaa Ser Asn Pro Leu Cys Pro Xaa Glu Thr Gln Glu Gly	
135 140 145	
TAACACTGTG GGTGCCCCCA CCTGTGCATT GGGACCACRA CTTACCCTC TTGGARACAA	594
TAAACTCTCA TGCCCCCAAA AAAAAAAAAA	623

## (2) INFORMATION FOR SEQ ID NO: 26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 1..16
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.1  
seq LVLTLCTLPLAVA/SA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met	Glu	Arg	Leu	Val	Leu	Thr	Leu	Cys	Thr	Leu	Pro	Leu	Ala	Val	Ala
1				5				10						15	

## (2) INFORMATION FOR SEQ ID NO: 27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 848 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 32..73
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.7  
seq LWLLFFLVTAIHA/EL

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

```

AACTTTGCCT TGTGTTTTCC ACCCTGAAAG A ATG TTG TGG CTG CTC TTT TTT CTG      55
                               Met Leu Trp Leu Leu Phe Phe Leu
                               -10

GTG ACT GCC ATT CAT GCT GAA CTC TGT CAA CCA GGT GCA GAA AAT GCT      103
Val Thr Ala Ile His Ala Glu Leu Cys Gln Pro Gly Ala Glu Asn Ala
-5                               1                               5                               10

TTT AAA GTG AGA CTT AGT ATC AGA ACA GCT CTG GGA GAT AAA GCA TAT      151
Phe Lys Val Arg Leu Ser Ile Arg Thr Ala Leu Gly Asp Lys Ala Tyr
15                               20                               25

GCC TGG GAT ACC AAT GAA GAA TAC CTC TTC AAA GCG ATG GTA GCT TTC      199
Ala Trp Asp Thr Asn Glu Glu Tyr Leu Phe Lys Ala Met Val Ala Phe
30                               35                               40

TCC ATG AGA AAA GTT CCC AAC AGA GAA GCA ACA GAA ATT TCC CAT GTC      247
Ser Met Arg Lys Val Pro Asn Arg Glu Ala Thr Glu Ile Ser His Val
45                               50                               55

CTA CTT TGC AAT GTA ACC CAG AGG GTA TCA TTC TGG TTT GTG GTT ACA      295
Leu Leu Cys Asn Val Thr Gln Arg Val Ser Phe Trp Phe Val Val Thr
60                               65                               70

GAC CCT TCA AAA AAT CAC ACC CTT CCT GCT GTT GAG GTG CAA TCA GCC      343
Asp Pro Ser Lys Asn His Thr Leu Pro Ala Val Glu Val Gln Ser Ala
75                               80                               85                               90

ATA AGA ATG AAC AAG AAC CGG ATC AAC AAT GCC TTC TTT CTA AAT GAC      391
Ile Arg Met Asn Lys Asn Arg Ile Asn Asn Ala Phe Phe Leu Asn Asp
95                               100                               105

CAA ACT CTG GAA TTT TTA AAA ATC CCT TCC ACA CTT GCA CCA CCC ATG      439
Gln Thr Leu Glu Phe Leu Lys Ile Pro Ser Thr Leu Ala Pro Pro Met
110                               115                               120

```



GAC CCA TCT GTG CCC ATC TGG ATT ATT ATA TTT GGT GTG ATA TTT TGC	487
Asp Pro Ser Val Pro Ile Trp Ile Ile Ile Phe Gly Val Ile Phe Cys	
125 130 135	
ATC ATC ATA GTT GCA ATT GCA CTA CTG ATT TTA TCA GGG ATC TGG CAA	535
Ile Ile Ile Val Ala Ile Ala Leu Leu Ile Leu Ser Gly Ile Trp Gln	
140 145 150	
CGT ADA ARA AAG AAC AAA GAA CCA TCT GAA GTG GAT GAC GCT GAA RAT	583
Arg Xaa Xaa Lys Asn Lys Glu Pro Ser Glu Val Asp Asp Ala Glu Xaa	
155 160 165 170	
AAK TGT GAA AAC ATG ATC ACA ATT GAA AAT GGC ATC CCC TCT GAT CCC	631
Xaa Cys Glu Asn Met Ile Thr Ile Glu Asn Gly Ile Pro Ser Asp Pro	
175 180 185	
CTG GAC ATG AAG GGA GGG CAT ATT AAT GAT GCC TTC ATG ACA GAG GAT	679
Leu Asp Met Lys Gly Gly His Ile Asn Asp Ala Phe Met Thr Glu Asp	
190 195 200	
GAG AGG CTC ACC CCT CTC TGAAGGGCTG TTGTTCTGCT TCCTCAARAA	727
Glu Arg Leu Thr Pro Leu	
205	
ATTAAACATT TGTTTCTGTG TGAAGGGCTG GCATCCTGAA ATACCAAGAG CAGATCATAT	787
WTTTGTGTTT ACCATTCTTC TTTTGTAATA AATTTTGAAT GTGCTTGAAA AAAAAAAAAA	847
C	848

## (2) INFORMATION FOR SEQ ID NO: 28:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 1..14
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.7  
seq LWLLFFLVTAIHA/EL

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Met	Leu	Trp	Leu	Leu	Phe	Phe	Leu	Val	Thr	Ala	Ile	His	Ala
1				5					10				

## (2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: NUCLEIC ACID
  - (C) STRANDEDNESS: SINGLE
  - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: Other nucleic acid
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GGGAAGATGG AGATAGTATT GCCTG

25

(2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: NUCLEIC ACID
  - (C) STRANDEDNESS: SINGLE
  - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: Other nucleic acid
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CTGCCATGTA CATGATAGAG AGATTC

26

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 546 base pairs
  - (B) TYPE: NUCLEIC ACID
  - (C) STRANDEDNESS: DOUBLE
  - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: Genomic DNA
- (ix) FEATURE:
  - (A) NAME/KEY: promoter
  - (B) LOCATION: 1..517
- (ix) FEATURE:
  - (A) NAME/KEY: transcription start site
  - (B) LOCATION: 518
- (ix) FEATURE:
  - (A) NAME/KEY: TF binding-site
  - (B) LOCATION: 17..25
  - (C) IDENTIFICATION METHOD: matinspector prediction
  - (D) OTHER INFORMATION: name CMYB\_01  
score 0.983  
sequence TGTCAGTTG
- (ix) FEATURE:
  - (A) NAME/KEY: TF binding-site

- (B) LOCATION: complement(18..27)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MYOD\_Q6  
score 0.961  
sequence CCCAACTGAC

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(75..85)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name S8\_01  
score 0.960  
sequence AATAGAATTAG

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 94..104
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name S8\_01  
score 0.966  
sequence AACTAAATTAG

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(129..139)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name DELTAEF1\_01  
score 0.960  
sequence GCACACCTCAG

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(155..165)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA\_C  
score 0.964  
sequence AGATAAATCCA

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 170..178
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name CMYB\_01  
score 0.958  
sequence CTTAGTTG

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 176..189
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA1\_02  
score 0.959  
sequence TTGTAGATAGGACA

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 180..190
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA\_C  
score 0.953  
sequence AGATAGGACAT

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 284..299
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name TAL1ALPHA47\_01  
score 0.973  
sequence CATAACAGATGGTAAG

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 284..299
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name TAL1BETA47\_01  
score 0.983  
sequence CATAACAGATGGTAAG

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 284..299
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name TAL1BETAITF2\_01  
score 0.978  
sequence CATAACAGATGGTAAG

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(287..296)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MYOD\_Q6  
score 0.954  
sequence ACCATCTGTT

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(302..314)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA1\_04  
score 0.953  
sequence TCAAGATAAAGTA

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 393..405
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name IK1\_01  
score 0.963  
sequence AGTTGGGAATTCC

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 393..404
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name IK2\_01  
score 0.985  
sequence AGTTGGGAATTC

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 396..405
- (C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name CREL\_01  
score 0.962  
sequence TGGGAATTCC

## (ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: 423..436  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name GATA1\_02  
score 0.950  
sequence TCAGTGATATGGCA

## (ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: complement(478..489)  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name SRY\_02  
score 0.951  
sequence TAAACAAAACA

## (ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: 486..493  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name E2F\_02  
score 0.957  
sequence TTAGCGC

## (ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: complement(514..521)  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name MZF1\_01  
score 0.975  
sequence TGAGGGGA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

```

TGAGTGCAGT GTTACATGTC AGTTGGGTTA AGTTTGTTAA TGTCATTCAA ATCTTCTATG   60
TCTTGATTTG CCTGCTAATT CTATTATTTT TGGAACATAA TTAGTTTGAT GGTTCTATTA  120
GTTATTGACT GAGGTGTGCT AATCTCCCAT TATGTGGATT TATCTATTTT TTCAGTTGTA  180
GATAGGACAT TGATAGATAC ATAAGTACCA GGACAAAAGC AGGGAGATCT TTTTCCAAA   240
ATCAGGAGAA AAAAATGACA TCTGGAAAAC CTATAGGGGA AGGCATAACA GATGGTAAGG   300
ATACTTTATC TTGAGTAGGA GAGCCTTCCT GTGGCAACGT GGAGAAGGGA AGAGGTCGTA  360
GAATTGAGGA GTCAGCTCAG TTAGAAGCAG GGAGTTGGGA ATTCCGTTCA TGTGATTTAG  420
CATCASTGAT ATGGCAAATG TGGGACTAAG GGTAGTGATC AGAGGGTTAA AATTGTGTGT  480
TTTGTTTTAG CGCTGCTGGG GCATCGCCTT GGGTCCCCTC AAACAGATTC CCATGAATCT  540
CTTCAT                                     546

```

## (2) INFORMATION FOR SEQ ID NO: 32:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GTACCAGGGA CTGTGACCAT TGC

23

## (2) INFORMATION FOR SEQ ID NO: 33:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CTGTGACCAT TGCTCCCAAG AGAG

24

## (2) INFORMATION FOR SEQ ID NO: 34:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 861 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Genomic DNA

## (ix) FEATURE:

- (A) NAME/KEY: promoter
- (B) LOCATION: 1..806

## (ix) FEATURE:

- (A) NAME/KEY: transcription start site
- (B) LOCATION: 807

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(60..70)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name NFY\_Q6  
score 0.956  
sequence GGACCAATCAT

## (ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: 70..77  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name MZF1\_01  
score 0.962  
sequence CCTGGGGA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: 124..132  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name CMYB\_01  
score 0.994  
sequence TGACCGTTG

(ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: complement(126..134)  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name VMYB\_02  
score 0.985  
sequence TCCAACGGT

(ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: 135..143  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name STAT\_01  
score 0.968  
sequence TTCCTGGAA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: complement(135..143)  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name STAT\_01  
score 0.951  
sequence TTCCAGGAA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: complement(252..259)  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name MZF1\_01  
score 0.956  
sequence TTGGGGGA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: 357..368  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name IK2\_01  
score 0.965  
sequence GAATGGGATTTC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site  
(B) LOCATION: 384..391  
(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name MZF1\_01  
score 0.986

sequence AGAGGGGA

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(410..421)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name SRY\_02  
score 0.955  
sequence GAAAACAAAACA

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 592..599
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MZF1\_01  
score 0.960  
sequence GAAGGGGA

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 618..627
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MYOD\_Q6  
score 0.981  
sequence AGCATCTGCC

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 632..642
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name DELTAEF1\_01  
score 0.958  
sequence TCCCACCTTC

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(813..823)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name S8\_01  
score 0.992  
sequence GAGGCAATTAT

## (ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(824..831)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MZF1\_01  
score 0.986  
sequence AGAGGGGA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

```
TACTATAGGG CACGCGTGGT CGACGGCCGG GCTGTTCTGG AGCAGAGGGC ATGTCAGTAA   60
TGATTGGTCC CTGGGGAAGG TCTGGCTGGC TCCAGCACAG TGAGGCATTT AGGTATCTCT   120
CGGTGACCGT TGGATTCCTG GAAGCAGTAG CTGTTCTGTT TGGATCTGGT AGGGACAGGG   180
CTCAGAGGGC TAGGCACGAG GGAAGGTCAG AGGAGAAGGS AGGSARGGCC CAGTGAGARG   240
```



GGAGCATGCC TTCCCCCAAC CCTGGCTTSC YCTTGGYMAM AGGGCGKTTY TGGGMACTTR 300  
AAYTCAGGGC CCAASCAGAA SCACAGGCC AKTCNTGGCT SMAAGCACAA TAGCCTGAAT 360  
GGGATTTTCAG GTTAGNCAGG GTGAGAGGGG AGGCTCTCTG GCTTAGTTTT GTTTTGTTTT 420  
CCAAATCAAG GTAACCTGCT CCCTTCTGCT ACGGGCCTTG GTCTTGGCTT GTCCTCACCC 480  
AGTCGGAACCT CCCTACCACT TTCAGGAGAG TGGTTTTAGG CCCGTGGGGC TGTCTGTTC 540  
CAAGCAGTGT GAGAACATGG CTGGTAGAGG CTCTAGCTGT GTGCGGGGCC TGAAGGGGAG 600  
TGGGTTCTCG CCCAAAGAGC ATCTGCCCAT TTCCCACCTT CCCTTCTCCC ACCAGAAGCT 660  
TGCCTGAGCT GTTTGGACAA AAATCCAAAC CCCACTTGGC TACTCTGGCC TGGCTTCAGC 720  
TTGGAACCCA ATACCTAGGC TTACAGGCCA TCCTGAGCCA GGGGCCTCTG GAAATTCTCT 780  
TCCTGATGGT CCTTTAGGTT TGGGCACAAA ATATAATTGC CTCTCCCCTC TCCCATTTTC 840  
TCTCTTGGGA GCAATGGTCA C 861

## (2) INFORMATION FOR SEQ ID NO: 35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CTGGGATGGA AGGCACGGTA

20

## (2) INFORMATION FOR SEQ ID NO: 36:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GAGACCACAC AGCTAGACAA

20

## (2) INFORMATION FOR SEQ ID NO: 37:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 555 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

- (A) NAME/KEY: promoter
- (B) LOCATION: 1..500

(ix) FEATURE:

- (A) NAME/KEY: transcription start site
- (B) LOCATION: 501

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 191..206
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name ARNT\_01  
score 0.964  
sequence GGA<sup>~</sup>CTCACGTGCTGCT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 193..204
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name NMYC\_01  
score 0.965  
sequence ACTCACGTGCTG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 193..204
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name USF\_01  
score 0.985  
sequence ACTCACGTGCTG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(193..204)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name USF\_01  
score 0.985  
sequence CAGCACGTGAGT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(193..204)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name NMYC\_01  
score 0.956  
sequence CAGCACGTGAGT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(193..204)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MYCMAX\_02  
score 0.972

sequence CAGCACGTGAGT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 195..202
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name USF\_C  
score 0.997  
sequence TCACGTGC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(195..202)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name USF\_C  
score 0.991  
sequence GCACGTGA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(210..217)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MZF1\_01  
score 0.968  
sequence CATGGGGA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 397..410
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name ELK1\_02  
score 0.963  
sequence CTCTCCGGAAGCCT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 400..409
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name CETS1P54\_01  
score 0.974  
sequence TCCGGAAGCC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(460..470)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name AP1\_Q4  
score 0.963  
sequence AGTGACTGAAC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(460..470)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name AP1FJ\_Q2  
score 0.961  
sequence AGTGACTGAAC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 547..555

(C) IDENTIFICATION METHOD: matinspector prediction  
(D) OTHER INFORMATION: name PADS\_C  
score 1.000  
sequence TGTGGTCTC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

```
CTATAGGGCA CGCKTGGTCG ACGGCCCCGGG CTGGTCTGGT CTGTKGTGGA GTCGGGTTGA    60
AGGACAGCAT TTGTKACATC TGGTCTACTG CACCTTCCCT CTGCCGTGCA CTTGGCCTTT    120
KAWAAGCTCA GCACCGGTGC CCATCACAGG GCCGGCAGCA CACACATCCC ATTACTCAGA    180
AGGAACTGAC GGA CTACGT GCTGCTCCGT CCCCATGAGC TCAGTGGACC TGTCTATGTA    240
GAGCAGTCAG ACAGTGCCTG GGATAGAGTG AGAGTTCAGC CAGTAAATCC AAGTGATTGT    300
CATTCCTGTC TGCATTAGTA ACTCCCAACC TAGATGTGAA AACTTAGTTC TTTCTCATAG    360
GTTGCTCTGC CCATGGTCCC ACTGCAGACC CAGGCACTCT CCGGAAGCCT GGAAATCACC    420
CGTGTCTTCT GCCTGCTCCC GCTCACATCC CAACTTGTG TTCAGTCACT GAGTTACAGA    480
TTTTGCCTCC TCAATTTCTC TTGTCTTAGT CCCATCCTCT GTTCCCCTGG CCAGTTTGTC    540
TAGCTGTGTG GTCTC                                                    555
```

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 247 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 152..193
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 11.8  
seq VLVALILLHSALA/QS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

```
AAAACCAGCG CCCCAGATTG AGGCGCGGGT TTGGTGGCGC GTTTCAGCGA AGTCGCACGT    60
GAAGGATAGC AGTGGCCTGA GAAAGACCCA GTCATGGCAG CCTCCAGCAT CAGTTCACCA    120
TGGRGGAAAG CATGTGTTCA AAGCCATTCT G ATG GTC CTA GTG GCC CTT ATC      172
                               Met Val Leu Val Ala Leu Ile
```

(2) INFORMATION FOR SEQ ID NO: 39:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

(2) INFORMATION FOR SEQ ID NO: 40:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 384 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 76..132
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10  
seq FLLLVTAAPRCILS/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

```

AATTTTTCCT TAAATTCAGG GTCCCGCTCA CATGGGAAAT ACTTTCTGAG AGTCCTCGAC      60
CTCCGGCGTA AGAAC ATG AAA GAC CTG TGG ATC TTC CTC CTC CTG GTG ACA      111
      Met Lys Asp Leu Trp Ile Phe Leu Leu Leu Val Thr
                        -15                      -10

GCT CCC AGA TGC ATC CTG TCC CAG GTG CAG CTG CAG GAG TCG GGC CCG      159
Ala Pro Arg Cys Ile Leu Ser Gln Val Gln Leu Gln Glu Ser Gly Pro
      -5                      1                      5

CGT CTA GTT AGG CCC TCG GAG ACC GTG TCC CTC AGC TGC ACC GTC TCC      207
Arg Leu Val Arg Pro Ser Glu Thr Val Ser Leu Ser Cys Thr Val Ser
      10                      15                      20                      25

GGT GAC TCC GTC AGC AGT GGT GAC CAT TAT TGG ACT TGG CTC CGG CAG      255
Gly Asp Ser Val Ser Ser Gly Asp His Tyr Trp Thr Trp Leu Arg Gln
                        30                      35                      40

CCC CCC GGG GGG GGA CTG GAG TGG ATT GGC TAT ATC TAC ACC ACT GGC      303
Pro Pro Gly Gly Gly Leu Glu Trp Ile Gly Tyr Ile Tyr Thr Thr Gly
                        45                      50                      55

AAA ATC GAC TAC AAC CCC TCS MTC AGG CGT CGA GTC ACC ATC TCC GTG      351
Lys Ile Asp Tyr Asn Pro Ser Xaa Arg Arg Arg Val Thr Ile Ser Val
      60                      65                      70

GAC ACC TCG AAG AAT CTT TTC TCC CTG ACA CGG      384
Asp Thr Ser Lys Asn Leu Phe Ser Leu Thr Arg
      75                      80

```

## (2) INFORMATION FOR SEQ ID NO: 41:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 296 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Placenta

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 60..122  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 10  
seq LLLCLQTWPEAAG/KD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

```

AAACTGACTC TGCTAGAACA GTGCCGTGCT TTTCCACAGA AGGTTAGACC CTGAAAGAG      59
ATG GCT CAG CAC CAC CTA TGG ATC TTG CTC CTT TGC CTG CAA ACC TGG      107
Met Ala Gln His His Leu Trp Ile Leu Leu Leu Cys Leu Gln Thr Trp
-20                               -15                               -10

CCG GAA GCA GCT GGA AAA GAC TCA GAA ATC TTC ACA GTG AAT GGG ATT      155
Pro Glu Ala Ala Gly Lys Asp Ser Glu Ile Phe Thr Val Asn Gly Ile
-5                               1                               5                               10

CTG GGA GAG TCA GTC ACT TTC CCT GTA AAT ATC CAA GAA CCA CGG CAA      203
Leu Gly Glu Ser Val Thr Phe Pro Val Asn Ile Gln Glu Pro Arg Gln
15                               20                               25

GTT AAA ATC ATT GCT TGG ACT TCT AAA ACA TCT GTT GCT TAT GTA ACA      251
Val Lys Ile Ile Ala Trp Thr Ser Lys Thr Ser Val Ala Tyr Val Thr
30                               35                               40

CCA GGA GAC TCA GAA ACA GCA CCC GTA GTT ACT GTG ACC CAC ATG      296
Pro Gly Asp Ser Glu Thr Ala Pro Val Val Thr Val Thr His Met
45                               50                               55

```

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 232 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 59..115  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 9.9  
seq FLFVVAATGVQS/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

```

ATCACATAAC ATCCACATCT CTCCTCTGAA GAAGGCCCTG GGAGCGCAST CANTCACC      58
ATG GAC TGG ACC TGG AGG TTC CTC TTT GTG GTG GCA GCA GCT ACA GGT      106
Met Asp Trp Thr Trp Arg Phe Leu Phe Val Val Ala Ala Ala Thr Gly
               -15                      -10                      -5
GTC CAG TCT CAG GTT CAA CTG GTG CAA TCT GGG GCT GAG GTG GTG AAG      154
Val Gln Ser Gln Val Gln Leu Val Gln Ser Gly Ala Glu Val Val Lys
               1                      5                      10
CCT GGG TCC TCG GTA AAG GTC TCC TGT AAG ACT TCT GGA GAC GGT TTC      202
Pro Gly Ser Ser Val Lys Val Ser Cys Lys Thr Ser Gly Asp Gly Phe
               15                      20                      25
AGC AAA TAT CCA ATC AAC TGG GTG CAA GGG      232
Ser Lys Tyr Pro Ile Asn Trp Val Gln Gly
               30                      35

```

## (2) INFORMATION FOR SEQ ID NO: 43:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 290 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 105..161
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.6  
seq GLLLLCLLPHRLA/LV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

```

ACACCATDTG CTTTAGTTTC AGTCCTCGTA TCATAGAAAA GTTCATGTCA TAAAGAAGT      60
TAAACCACT CTTGAATAAT TGGAACCCCTT ATGCCAATTG TCTA ATG TCA ATT TGT      116
Met Ser Ile Cys
TTT CTG GGA TTG CTT CTT CTA TGT CTT CTT CCT CAT CGT CTG GCA CTG      164
Phe Leu Gly Leu Leu Leu Leu Cys Leu Leu Pro His Arg Leu Ala Leu
-15                      -10                      -5                      1
GTT CAG AAA CAC TCA TCT CCA TCA AGT CGT CTT CTG CTA ATT CCT GTG      212
Val Gln Lys His Ser Ser Pro Ser Ser Arg Leu Leu Leu Ile Pro Val
               5                      10                      15

```



GTG CAG TGT CTA CTA GCT CTT GAA TTT CTC CAA GAT CCA TAT CTT GAT 260  
Val Gln Cys Leu Leu Ala Leu Glu Phe Leu Gln Asp Pro Tyr Leu Asp  
20 25 30

ATC TTC AAC CTG CCC CTG CCC CCA CCT TGG 290  
Ile Phe Asn Leu Pro Leu Pro Pro Pro Trp  
35 40

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 148..198  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 9.2  
seq LVLLILPLSSLS/KV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

ATTTCTCTAA	ACAGCTTGTT	ACATGTTTCT	TTTAAGTTCA	GTCAAATTGC	ATAGGAACCT		60
ATTTTTAAGT	AGAAGATAGA	TTCTGAAGCT	TTATGTATAC	GGATTAAAAA	TTGAGGCTAA		120
AATTTGATCT	TATTTTAACT	TTATAAT	ATG ATA GGA TTT CTA GTG CTT TTA ATA				174
			Met Ile Gly Phe Leu Val Leu Leu Ile				
			-15		-10		
CTT CCT CTC CTT TCT TCC TTA TCC AAA GTC AGC TCC AAG							213
Leu Pro Leu Leu Ser Ser Leu Ser Lys Val Ser Ser Lys							
	-5		1		5		

(2) INFORMATION FOR SEO ID NO: 45:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganqlia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 59..151
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.9  
seq LLMSLLVSTVTWQ/IS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

```

ATTTTTTTTAA CATTGTCAAG CTTTATATCA TGTTATTTCC CAAGTATAAT TCACATGG      58
ATG CAG TGC TTA CTT TCA GTC CTT ATG GCA CAG TTT ATT KCT CAT TTC      106
Met Gln Cys Leu Leu Ser Val Leu Met Ala Gln Phe Ile Xaa His Phe
-30                      -25                      -20

TTA TCT TTA TTA ATG TCT TTG TTG GTG AGT ACA GTC ACT TGG CAG ATC      154
Leu Ser Leu Leu Met Ser Leu Leu Val Ser Thr Val Thr Trp Gln Ile
-15                      -10                      -5                      1

TCC CGC ACC CCC TGG CAT GGG      175
Ser Arg Thr Pro Trp His Gly
5

```

## (2) INFORMATION FOR SEQ ID NO: 46:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 341 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 81..137
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.9  
seq WIFFLATLKGVC/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

```

ACCTCTGGGA GAGGAGCCCC AGCCCTGAGA TCCCCGGGTG TTTCCATTCA GCGACCAGCA      60
CTGAAACACAG AGGACTCACC ATG GAG TTG GGA CTG AGC TGG ATC TTC TTC TTG      113
Met Glu Leu Gly Leu Ser Trp Ile Phe Phe Leu
-15                      -10

GCT ACT TTA AAA GGT GTC CAA TGT CAG GTG AGG CTG CTG GAG TCT GCG      161
Ala Thr Leu Lys Gly Val Gln Cys Gln Val Arg Leu Leu Glu Ser Ala
-5                      1                      5

```

GGA GGC CTG CAA GAG CCT GGC GGG GCC CTG AGA CTC TCC TGT GCA GTC	209
Gly Gly Leu Gln Glu Pro Gly Gly Ala Leu Arg Leu Ser Cys Ala Val	
10 15 20	
AGT GGC TTC ATC TTC AAT GAC TTC GCC ATG CAT TGG GTC CGC CAG ACG	257
Ser Gly Phe Ile Phe Asn Asp Phe Ala Met His Trp Val Arg Gln Thr	
25 30 35 40	
CCA GGG AAG GGC CTG GAG TGG GTC GCA GGC ATA AAT TGG GAT GGC HWC	305
Pro Gly Lys Gly Leu Glu Trp Val Ala Gly Ile Asn Trp Asp Gly Xaa	
45 50 55	
ATT TTA GGG TAT GCG GAC TCT GTG AAG GGC CGC AGG	341
Ile Leu Gly Tyr Ala Asp Ser Val Lys Gly Arg Arg	
60 65	

## (2) INFORMATION FOR SEQ ID NO: 47:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 195 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 19..63
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.6  
seq SVSLALLSGWVGS/RQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

ACCCTTTCCC TGTTAGAC ATG GTA AGT GTG AGT TTA GCG CTG CTG TCC GGA	51
Met Val Ser Val Ser Leu Ala Leu Leu Ser Gly	
-15 -10 -5	
TGG GTT GGT AGC AGA CAG GGT AGA GTA GGG TTA AGC ACA CTG GTC ACC	99
Trp Val Gly Ser Arg Gln Gly Arg Val Gly Leu Ser Thr Leu Val Thr	
1 5 10	
TTA GGA TTG GTT TCC TGG TGC TGG AGA ATG GTT AGG ACA CAG GCC TTG	147
Leu Gly Leu Val Ser Trp Cys Trp Arg Met Val Arg Thr Gln Ala Leu	
15 20 25	
GAA GST TTT TTG AGT GTG AAA TAT TAC TCA GCG TTT TCT GCA GAC CAG	195
Glu Gly Phe Leu Ser Val Lys Tyr Tyr Ser Ala Phe Ser Ala Asp Gln	
30 35 40	

## (2) INFORMATION FOR SEQ ID NO: 48:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 342 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Placenta

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 172..234
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.5  
seq LPLLLSWVAGGFG/NA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

```

AGCGGTGCCT GGCCTCCCCT CCCAGACTGC AGGGACAGCA CCCGGTAACT GCGAGTGGAG      60
CGGAGGACCC GAGCGGCTGA GGAGAGAGGA GGCGGCGGCT TAGCTGCTAC GGGGTCCGGC      120
CGGCGCCCTC CCGAGGGGGG CTCAGGAGGA GGAAGGAGGA CCCGTGCGAG A ATG CCT      177
                                   Met Pro
                                   -20

CTG CCC TGG AGC CTT GCG CTC CCG CTG CTG CTC TCC TGG GTG GCA GGT      225
Leu Pro Trp Ser Leu Ala Leu Pro Leu Leu Leu Ser Trp Val Ala Gly
                                   -15                               -5

GGT TTC GGG AAC GCG GCC AGT GCA AGG CAT CAC GGG TTG TTA GCA TCG      273
Gly Phe Gly Asn Ala Ala Ser Ala Arg His His Gly Leu Leu Ala Ser
                                   1                               5                               10

GCA CGT CAG CCT GGG GTC TGT CAC TAT GGA ACT AAA CTG GCC TGC TGC      321
Ala Arg Gln Pro Gly Val Cys His Tyr Gly Thr Lys Leu Ala Cys Cys
                                   15                               20                               25

TAC GGC TGG AGA AGA AAC AGC      342
Tyr Gly Trp Arg Arg Asn Ser
                                   30                               35

```

## (2) INFORMATION FOR SEQ ID NO: 49:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 377 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 36..275
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.4  
seq FVVFSFLFLICAMA/GD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

```

AGCAGAAAAC TAACTGAAAA ACGAGAACCT ACTGT ATG GTT AGT AAT TTC TTC      53
                               Met Val Ser Asn Phe Phe
                               -80                      -75

CAT GTC ATA CAA GTA TTC GAG AAA TCT GCT ACC TTG ATT AGT AAG ACT      101
His Val Ile Gln Val Phe Glu Lys Ser Ala Thr Leu Ile Ser Lys Thr
                               -70                      -65                      -60

GAA CAC ATT GGT TTT GTC ATT TAT TCA TGG ASG AAG TCC ACC ACC CAC      149
Glu His Ile Gly Phe Val Ile Tyr Ser Trp Xaa Lys Ser Thr Thr His
                               -55                      -50                      -45

TTG GGG AGC AGA AGG AAA TTT GCC ATC TCA ATT TAC TTA TCA GAA GTT      197
Leu Gly Ser Arg Arg Lys Phe Ala Ile Ser Ile Tyr Leu Ser Glu Val
                               -40                      -35                      -30

TCT TTG CAG AAA TAT GAT TGT CCC TTC AGT GGG ACA TCA TTT GTG GTC      245
Ser Leu Gln Lys Tyr Asp Cys Pro Phe Ser Gly Thr Ser Phe Val Val
                               -25                      -20                      -15

TTC TCT CTC TTT TTG ATC TGT GCA ATG GCT GGA GAT GTA GTC TAC GCT      293
Phe Ser Leu Phe Leu Ile Cys Ala Met Ala Gly Asp Val Val Tyr Ala
-10                      -5                      1                      5

GAC ATC AAA ACT GTT CGG ACT TCC CCG TTA GAA CTC GCG TYT TTN CTT      341
Asp Ile Lys Thr Val Arg Thr Ser Pro Leu Glu Leu Ala Xaa Xaa Leu
                               10                      15                      20

CAG AGA TCT GKG SCT TTC AAC TTT TCT AGB MTC CGG      377
Gln Arg Ser Xaa Xaa Phe Asn Phe Ser Xaa Xaa Arg
                               25                      30

```

## (2) INFORMATION FOR SEQ ID NO: 50:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 1..132  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 8.2  
 seq ICLACVLFPLLRT/SD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

ATG CGA CNN TTT TGG TTT CTC ATG TAC CCC TTT CGC TTC CAT GAC TGC	48
Met Arg Xaa Phe Trp Phe Leu Met Tyr Pro Phe Arg Phe His Asp Cys	
-40 -35 -30	
AAA CAG AAA TAT GAC CTG TAC ATC AGC ATT GCT GGC TGG CTG ATC ATC	96
Lys Gln Lys Tyr Asp Leu Tyr Ile Ser Ile Ala Gly Trp Leu Ile Ile	
-25 -20 -15	
TGC CTT GCC TGT GTA CTC TTT CCA CTC CTC AGA ACC AGT GAT GAT ACC	144
Cys Leu Ala Cys Val Leu Phe Pro Leu Leu Arg Thr Ser Asp Asp Thr	
-10 -5 1	
CCT GGC AAT AGG ACC AAA TGC TTT GTG GAT CTT CCT ACC AGG AAT GTC	192
Pro Gly Asn Arg Thr Lys Cys Phe Val Asp Leu Pro Thr Arg Asn Val	
5 10 15 20	
AAC CTG GCC CAG TCC GTT GTT ATG ATG ACC ATT GGC GAG TTG ATT GGG	240
Asn Leu Ala Gln Ser Val Val Met Met Thr Ile Gly Glu Leu Ile Gly	
25 30 35	

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 193 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 2..52  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 7  
 seq CCLFTCFIPICIS/CK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

C ATG GTT TCC TTA TGT TGT CTT TTC ACT TGT TTC TTT ATT CCT TGT ATT	49
Met Val Ser Leu Cys Cys Leu Phe Thr Cys Phe Phe Pro Cys Ile	
-15 -10 -5	
TCC TGT AAG CTA GAA ATG TGG GGA CTC GAT GAG CCT AAA GTT AAA CCA	97
Ser Cys Lys Leu Glu Met Trp Gly Leu Asp Glu Pro Lys Val Lys Pro	

1	5	10	15	
TTC TGG CAA GAA TGC GTC CTA GGG GAT GTT GTG GGC NTC ATC CTG CAG				145
Phe Trp Gln Glu Cys Val Leu Gly Asp Val Val Gly Xaa Ile Leu Gln				
	20	25	30	
CAC AGG AGG CAG CCC CCG GTG CCA CGT TCC ATA CTA GTA ATG GGG GCC				193
His Arg Arg Gln Pro Pro Val Pro Arg Ser Ile Leu Val Met Gly Ala				
	35	40	45	

## (2) INFORMATION FOR SEQ ID NO: 52:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 146 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Placenta

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 57..137
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.9  
seq ILLLVITYSPIAYS/HS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

AGCAGGCAGC CTGTGCCTTT TCTCCAAACG AAGACAGCAC TTTGAAAATT CTTTCA ATG	59
	Met
GAT TTT TTT TTC CTT GAA AGA TCG TAC TGG GGG AAA ATG ATA CTT CTA	107
Asp Phe Phe Phe Leu Glu Arg Ser Tyr Trp Gly Lys Met Ile Leu Leu	
-23 -20 -15	
TTA GTT ACA TAT TCT CCA ATC GCA TAC TCG CAC TCC CGG	146
Leu Val Thr Tyr Ser Pro Ile Ala Tyr Ser His Ser Arg	
-10 -5 1	

## (2) INFORMATION FOR SEQ ID NO: 53:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 334 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 170..247  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.9  
seq LWVLLLLCAHVVT/LV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

```
ATTTTTCCTG TGGTGGGTTC ACACGCAGCT AGACACAGCT AACTTGAGTC TTGGAGCTCC    60
TAGAGGGAAG CTTCTGGAAA GGAAGGCTCT TCAGGACCTC TTAGGAGCCA GAGAAGAGGA    120
CGTTGTCACA GATAAAGAGC CAGGCTCACC AGCTCCTGAC GCATGCATC ATG ACC ATG    178
                                     Met Thr Met
                                     -25

AGA CAC AAC TGG ACA CCA GAC CTC AGC CCT TTG TGG GTC CTG CTC CTG    226
Arg His Asn Trp Thr Pro Asp Leu Ser Pro Leu Trp Val Leu Leu Leu
      -20                      -15                      -10

TGT GCC CAC GTC GTC ACT CTC CTG GTC AGA GCC ACA CCT GTC TCG CAG    274
Cys Ala His Val Val Thr Leu Leu Val Arg Ala Thr Pro Val Ser Gln
      -5                      1                      5

ACC AYC ACA GCT GCS ACT GCC TCA GTT AGA AGC ACA AAG GAC CCC TGC    322
Thr Xaa Thr Ala Ala Thr Ala Ser Val Arg Ser Thr Lys Asp Pro Cys
    10                      15                      20                      25

CCC ACC CAG GGG    334
Pro Thr Gln Gly
```

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 254 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 78..221  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.8  
seq ILRMLLSLQPVLQ/DA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:



```

ATAGCAATTG GGTACGTGTT GAAGAGCGTG ACTGTTGCAA TGA CTGCTAC CTTGCATTAG      60
AACATGGGCG TCAGTTC ATG GAT AAC ATG TCA GGA GGA AAA GTT GAT GAA      110
           Met Asp Asn Met Ser Gly Gly Lys Val Asp Glu
           -45                               -40

GCA CTT GTG AAA AGT TCA TGC TTA CAC CCC TGG TCC AAA AGA AAC GAT      158
Ala Leu Val Lys Ser Ser Cys Leu His Pro Trp Ser Lys Arg Asn Asp
      -35                               -30                               -25

GTG AGT ATG CAG TGC TCA CAG GAT ATA CTT CGA ATG CTC CTC TCT CTT      206
Val Ser Met Gln Cys Ser Gln Asp Ile Leu Arg Met Leu Leu Ser Leu
      -20                               -15                               -10

CAG CCA GTT CTT CAG GAT GCC ATT CAG AAA AAA AGA ACA GTA AGA CAG      254
Gln Pro Val Leu Gln Asp Ala Ile Gln Lys Lys Arg Thr Val Arg Gln
      -5                               1                               5                               10

```

## (2) INFORMATION FOR SEQ ID NO: 55:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 181 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 5..139
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.7  
seq AALVLWTLPGAQR/RG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

```

AAGG ATG YGA CTG CAA GGC CAG GAA GCT ACA GGG AAA GTT CTG ATC AAA      49
Met Xaa Leu Gln Gly Gln Glu Ala Thr Gly Lys Val Leu Ile Lys
-45                               -40                               -35

ATA CAC AAA GAC ACA AGC CAG GTC CCC ACC GCG CKW GGC GAT GCA TCC      97
Ile His Lys Asp Thr Ser Gln Val Pro Thr Ala Xaa Gly Asp Ala Ser
-30                               -25                               -20                               -15

ATA GCA GCC TTG GTG CTG TGG ACA CTC CCT GGG GCC CAG CGA AGG GGA      145
Ile Ala Ala Leu Val Leu Trp Thr Leu Pro Gly Ala Gln Arg Arg Gly
      -10                               -5                               1

GAG TTT GCT CCC AAA GGC GCA CCA ATG ACC AAC AGG      181
Glu Phe Ala Pro Lys Gly Ala Pro Met Thr Asn Arg
      5                               10

```

## (2) INFORMATION FOR SEQ ID NO: 56:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 22..96
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.6  
seq ILVLILFPTSCVM/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

```
AACCTACACA ATGATAGTGT A ATG ACT GAG CAC TCA CTG ACG CAT CAA GGG      51
                  Met Thr Glu His Ser Leu Thr His Gln Gly
                  -25                      -20

ATC CCA ATT CTA GTC TTG ATT CTA TTT CCA ACT AGT TGT GTC ATG CAA      99
Ile Pro Ile Leu Val Leu Ile Leu Phe Pro Thr Ser Cys Val Met Gln
-15                      -10                      -5                      1

GTC CTC TGG
Val Leu Trp                                         108
```

## (2) INFORMATION FOR SEQ ID NO: 57:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 118..174
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.5  
seq RFIFLTSLQLISS/SY

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

```

AAATCTTGAA GTGTATTGTT GAATCATAGA TGTTTTTATG TGGCTATCTA TTCCTTACTT      60
AATCCTTCAT TTTCAAGAAG ATTGTGTCGT WAAAGCCCTT AAAATTTGGT ACCCTTT      117
ATG TAC ATT GGT GGT CTG AGA TTC ATT TTT CTC ACC TCT TTA CAA CTA      165
Met Tyr Ile Gly Gly Leu Arg Phe Ile Phe Leu Thr Ser Leu Gln Leu
               -15                      -10                      -5
ATT TCA AGC AGC TAT GTT ACC ACT TTA TTA AAA AAA AAC ACA CTT AGG      213
Ile Ser Ser Ser Tyr Val Thr Thr Leu Leu Lys Lys Asn Thr Leu Arg
               1                      5                      10

```

## (2) INFORMATION FOR SEQ ID NO: 58:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 227 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 105..218
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.5  
seq LIFFSLIFLNLFA/IS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

```

ATGATTAAAG AAATGATCTT TTAAAGCAAA ATTGTTTGCT GTAGCCAGTG ACAGCTTATT      60
TAAAGAAAGT GTTAACTTA TTTGATTTTA GCATGTTTAA GTAA ATG TCT GTT AGT      116
               Met Ser Val Ser
               -35
CTT AAA CAC ATT CAC TTG CAT TTT ATT ATT ATG TCG GTA CTT GTA TTT      164
Leu Lys His Ile His Leu His Phe Ile Ile Met Ser Val Leu Val Phe
               -30                      -25                      -20
TGG AAC TGT AGT CAT TTG ATT TTC TTT TCC TTG ATT TTT TTA AAC CTG      212
Trp Asn Cys Ser His Leu Ile Phe Phe Ser Leu Ile Phe Leu Asn Leu
               -15                      -10                      -5
TTT GCG ATC TCC TGG      227
Phe Ala Ile Ser Trp
               1

```

## (2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 186 base pairs  
    (B) TYPE: NUCLEIC ACID  
    (C) STRANDEDNESS: DOUBLE  
    (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:  
    (A) ORGANISM: Homo Sapiens  
    (F) TISSUE TYPE: Lymph ganglia
- (ix) FEATURE:  
    (A) NAME/KEY: sig\_peptide  
    (B) LOCATION: 106..171  
    (C) IDENTIFICATION METHOD: Von Heijne matrix  
    (D) OTHER INFORMATION: score 6.3  
                                seq MVSFLSXPFLCSA/KP
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

```
AAGCYCMGNY ATCCTTGGAATAAACCTCAA AARGCCTTAG GCTAATGCAC AGGGTTCTTC      60
CTTCTAGTTC TGA CTCACCT TCCTGACCTC ATTTTATGCC TCACT ATG  ARD  ASA  CTT      117
                                   Met Xaa Xaa Leu
                                   -20

GGT ADT TKT AGA TTT ATG GTC TCT TTC CTA TCA MWC CCC TTT CTT TGT      165
Gly Xaa Xaa Arg Phe Met Val Ser Phe Leu Ser Xaa Pro Phe Leu Cys
      -15                      -10                      -5

TCA GCA AAA CCA TCC TCC GGG      186
Ser Ala Lys Pro Ser Ser Gly
      1                      5
```

(2) INFORMATION FOR SEQ ID NO: 60:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 267 base pairs  
    (B) TYPE: NUCLEIC ACID  
    (C) STRANDEDNESS: DOUBLE  
    (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:  
    (A) ORGANISM: Homo Sapiens  
    (F) TISSUE TYPE: Lymph ganglia
- (ix) FEATURE:  
    (A) NAME/KEY: sig\_peptide  
    (B) LOCATION: 55..111  
    (C) IDENTIFICATION METHOD: Von Heijne matrix  
    (D) OTHER INFORMATION: score 6.3  
                                seq ILVSVAAATGAHS/QL
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

```

ACCAAHNYAC CACACCCCTC CTTGGGAGAA TCCCCTGGAT CACAGCTCCT CACC ATG      57
                                     Met

GAC TGG ACC TGG TAC ATC CTT GTC TCG GTG GCA GCA GCA ACA GGT GCC      105
Asp Trp Thr Trp Tyr Ile Leu Val Ser Val Ala Ala Ala Thr Gly Ala
      -15                      -10                      -5

CAC TCC CAA CTT CAG CTG CTG CAG TCT GGA AGT GAC ATC AAG AAG CCT      153
His Ser Gln Leu Gln Leu Leu Gln Ser Gly Ser Asp Ile Lys Lys Pro
      1                      5                      10

GGG GCC TCA ATG AAC GTC TCC TGC AAG GCA TCT GGC GGC AGT ATT AGC      201
Gly Ala Ser Met Asn Val Ser Cys Lys Ala Ser Gly Gly Ser Ile Ser
      15                      20                      25                      30

ACC CGT GGA ATT AGT TGG GTC CGA CAG GTC CCT GGA CAG GGG CTT GAG      249
Thr Arg Gly Ile Ser Trp Val Arg Gln Val Pro Gly Gln Gly Leu Glu
      35                      40                      45

TGG ATG GGA TGG ATC GGG
Trp Met Gly Trp Ile Gly
      50

```

## (2) INFORMATION FOR SEQ ID NO: 61:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 259 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 92..226
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.3  
seq VACVLSSLIAVNS/AH

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

```

AAAGCAGYST ACCTTGATG GGCAGTGGAG CAAGCTAAGG AGACACAGGA GATATATGGG      60

AACCTAAAGT GGAATTTTCA AATTCTTGCT C ATG ATT TCA AAA TTC TCT TCA      112
Met Ile Ser Lys Phe Ser
      -45                      -40

AAG GCG TAC TCT GTA AGG GGC TTA GAG CTC TTC TCC CTA TTG CCT ATT      160
Lys Ala Tyr Ser Val Arg Gly Leu Glu Leu Phe Ser Leu Leu Pro Ile
      -35                      -30                      -25

AAC CCA TCC CCA AAC AGT GCC ATT NNG GTA GCT TGT GTG CTC TCT TCT      208

```

Asn	Pro	Ser	Pro	Asn	Ser	Ala	Ile	Xaa	Val	Ala	Cys	Val	Leu	Ser	Ser		
	-20						-15					-10					
CTT	ATT	GCT	GTT	AAC	TCA	GCT	CAC	CCA	GAA	AGT	ACT	ATT	GAC	ACC	CGC	256	
Leu	Ile	Ala	Val	Asn	Ser	Ala	His	Pro	Glu	Ser	Thr	Ile	Asp	Thr	Arg		
	-5					1				5					10		
TGG																259	
Trp																	

## (2) INFORMATION FOR SEQ ID NO: 62:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 165 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 76..159
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.3  
seq LLFLIFSLNLRG/VG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

AGATGTCCAT	AATTATTGGT	AACTCAGTTA	CCTTCTAACT	AATAGGCTGG	TTCAGGAGAC	60	
TCTCCCAGTT	TATAA	ATG GTT CTC	TTG GGA GCC	TTT GGA AGC	TGT ATT AAA	111	
	Met	Val Leu	Leu Gly	Ala Phe	Gly Ser	Cys Ile	Lys
		-25			-20		
TCT TTC	AGT CTT	TTA TTT	CTA ATT	TTT TCT	CTT AAT	CTA AAT	AGA GGC
Ser Phe	Ser Leu	Leu Phe	Leu Ile	Phe Ser	Leu Asn	Leu Asn	Arg Gly
-15		-10			-5		
GTC GGG							165
Val Gly							
1							

## (2) INFORMATION FOR SEQ ID NO: 63:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 388 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 74..178  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.3  
seq ALKLLSPGXSGS/SS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

```

AACTTAACAT GCGGCGGCG GCGCACTGCC GAGGCGCCTG AGCGGGTCGC GAGCGTGGTG      60
TTACTCTCCA GTC ATG GCG GCC CGA CAG GCC GTG GGC AGC GGG GCT CAG      109
      Met Ala Ala Arg Gln Ala Val Gly Ser Gly Ala Gln
      -35                      -30                      -25

GAG ACA TGC GGT CTG GAT CGG ATT TTG GAG GCA TTG AAG CTG CTG CTG      157
Glu Thr Cys Gly Leu Asp Arg Ile Leu Glu Ala Leu Lys Leu Leu Leu
      -20                      -15                      -10

AGC CCG GGM SGC TCG GGC TCA AGT TCA CTA CAG GTC ACA AAA CAT GAT      205
Ser Pro Gly Xaa Ser Gly Ser Ser Ser Leu Gln Val Thr Lys His Asp
      -5                      1                      5

GTC TTG TTG GCT ACT TTA AAA TCT AAC CTG TCT GCT TTG GAG GAC AAG      253
Val Leu Leu Ala Thr Leu Lys Ser Asn Leu Ser Ala Leu Glu Asp Lys
      10                      15                      20                      25

TTT CTG AAG GAT CCT CAG TGG AAG AAT CTG AAA CTC CTA AGA GAT GAA      301
Phe Leu Lys Asp Pro Gln Trp Lys Asn Leu Lys Leu Leu Arg Asp Glu
      30                      35                      40

ATT GCT GAT AAG GCA GAA TGG CCA CAA AAC TCT GTG GAT GTC ACT TGG      349
Ile Ala Asp Lys Ala Glu Trp Pro Gln Asn Ser Val Asp Val Thr Trp
      45                      50                      55

AGT TTT ACC TCT CAA ACC TTG TTG TTG CTT TTG TGC TTG      388
Ser Phe Thr Ser Gln Thr Leu Leu Leu Leu Leu Cys Leu
      60                      65                      70

```

## (2) INFORMATION FOR SEQ ID NO: 64:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 118..174
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.2  
seq LALFLMALGFSCI/HK

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

```

ATTGTAAYAA AAATATTTAC TACTCTTACT GGACTTAGTT TCCTTTTTGT GGTTCCTGTT      60
ACGTGGTCTA TAAGTTTCAA GCAGTGAGCT TAATTTGTGC GTCATAAAAA TTTGGTT      117
ATG AGT ACA CAG AAG GGA CTT GCT CTG TTT CTC ATG GCC CTT GGC TTT      165
Met Ser Thr Gln Lys Gly Leu Ala Leu Phe Leu Met Ala Leu Gly Phe
               -15                      -10                      -5
TCA TGC ATA CAC AAG AAG TTT CAG GAG TCA GAG GAG GGT AAG CAC CAT      213
Ser Cys Ile His Lys Lys Phe Gln Glu Ser Glu Glu Gly Lys His His
               1                      5                      10
ATG GGT GGA ATT AAT AGG TCT CAT TGG GTT AAG TCT CGA AAG AGC TGT      261
Met Gly Gly Ile Asn Arg Ser His Trp Val Lys Ser Arg Lys Ser Cys
               15                      20                      25
TTA ATA AAT AGC CAA CGC AAG
Leu Ile Asn Ser Gln Arg Lys
               30                      35

```

## (2) INFORMATION FOR SEQ ID NO: 65:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 147 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 67..141
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.1  
seq YFLIVFFVFLCNC/HQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

```

ACAGACATTT CATACTGGCC ATCTTGATGT AGATTACATT GCCATCCTGT GTCCTGATTG      60
TACTTT ATG AAA GAT GTA GAA ATA ATC ATG ATA TTT CAC GGT TAT TTC      108
Met Lys Asp Val Glu Ile Ile Met Ile Phe His Gly Tyr Phe
      -25                      -20                      -15

```



TTG ATT GTG TTT TTT GTG TTC CTA TGC AAC TGC CAC CAG  
 Leu Ile Val Phe Phe Val Phe Leu Cys Asn Cys His Gln  
 -10 -5 1

147

## (2) INFORMATION FOR SEQ ID NO: 66:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 345 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 214..339
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.1  
seq AILLLQSQCAYWA/LP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

AARTTGAGCT TGGGGACTGC AGCTGTGGGG AGATTTCAGT GCATTGCCTC CCCTGGGTGC 60  
 TCTTCATCTT GGATTGAAA GTTGAGAGCA GCATGTTTTG CCCACTGAAA CTCATCCTGS 120  
 TGRSAGTGTA MTGGATTATT CCTTGGGCCT GAATGACTTG AATGTTTCCC CGCCTGAGCT 180  
 AACAGTCCAT GTGGGTGATT CAGCTCTGAT GGG ATG TGT TTT CCA GAG CAC AGA 234  
 Met Cys Phe Pro Glu His Arg  
 -40  
 AGA CAA ATG TAT ATT CAA GAT AGA CTG GAC TCT GTC ACC AGG AGA GCA 282  
 Arg Gln Met Tyr Ile Gln Asp Arg Leu Asp Ser Val Thr Arg Arg Ala  
 -35 -30 -25 -20  
 CGC CAA GGA CGA ATA TGT GCT ATA CTA TTA CTC CAA TCT CAG TGT GCC 330  
 Arg Gln Gly Arg Ile Cys Ala Ile Leu Leu Leu Gln Ser Gln Cys Ala  
 -15 -10 -5  
 TAT TGG GCG CTT CCA 345  
 Tyr Trp Ala Leu Pro  
 1

## (2) INFORMATION FOR SEQ ID NO: 67:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 253 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: 155..223

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 6.1  
seq SSILSTFVSWLSA/FY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

```

AATATAATTT GATATATAAA TATACTTCCC CACGCAATTA ATAAGCAAGT GCTGGAGAGA    60
CATTTTAAGA TTCTGTAAAT ATCCTGCTGC ACATGAAAAG TTTCCCCCTA GATTGAGCGT   120
CTGCTGATGA TTCTTGTGTG ATCCAGCATT TACT ATG TTG GTT GTA AAA CAA TGC   175
                               Met Leu Val Val Lys Gln Cys
                               -20

TTT TCT GAC TCC AGT ATT CTC TCC ACA TTC GTA AGT TGG CTC TCA GCA    223
Phe Ser Asp Ser Ser Ile Leu Ser Thr Phe Val Ser Trp Leu Ser Ala
-15                               -10                               -5

TTC TAC TGT AAA GAA GGA CCC TCC TCG GGG                                253
Phe Tyr Cys Lys Glu Gly Pro Ser Ser Gly
 1                5                10

```

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 350 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: 39..134

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 6.1  
seq LPLLTSALHGLQQ/QH

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

```

AGCCCCAGAT CCTGAAGGAG GTGCAGAGCC CAGAGGGG ATG ATC KCG CTG AGG GAC    56

```

Met Ile Xaa Leu Arg Asp  
-30

ACA GCT GCC TCC CTC CGC CTT GAG AGA GAC ACA AGG CAG TTG CCA CTG	104
Thr Ala Ala Ser Leu Arg Leu Glu Arg Asp Thr Arg Gln Leu Pro Leu	
-25 -20 -15	
CTC ACC AGT GCC CTG CAC GGA CTG CAG CAG CAG CAC CCA GCC TTC TCT	152
Leu Thr Ser Ala Leu His Gly Leu Gln Gln Gln His Pro Ala Phe Ser	
-10 -5 1 5	
GGT GTG GCA CGG CTG GCC AAG CGG TGG GTG CGT GCC CAG CTT CTT GGT	200
Gly Val Ala Arg Leu Ala Lys Arg Trp Val Arg Ala Gln Leu Leu Gly	
10 15 20	
GAG GGT TTC GCT GAT GAG AGC CTG GAT CTG GTG GCC GCT GCC CTT TTC	248
Glu Gly Phe Ala Asp Glu Ser Leu Asp Leu Val Ala Ala Ala Leu Phe	
25 30 35	
CTG CAC CCT GAG CCC TTC ACC CCT CCG AGT TCC CCC CAG GTT GGC TTC	296
Leu His Pro Glu Pro Phe Thr Pro Pro Ser Ser Pro Gln Val Gly Phe	
40 45 50	
CTT CGA TTC CTT TTC TTG GTA TCA ACG TTT GAT TGG AAG AAC AAC CCC	344
Leu Arg Phe Leu Phe Leu Val Ser Thr Phe Asp Trp Lys Asn Asn Pro	
55 60 65 70	
CTC GGG	350
Leu Gly	

## (2) INFORMATION FOR SEQ ID NO: 69:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 315 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 265..306
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9  
seq ITMMLALISVCLF/AF

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

AAATAATTGA AGTCATACCA GCTTGACTTA GATAACATAC AAAATTCTGT TCCTTTTTTCA	60
CTGTGTTGATG TAACAAAATT ACATCTTTAT ACATTGTGTA CACAACAACA TAAGCTAATA	120
AAATAATACA CTCATTTCTT AAATTGTAGA AAACAAAATT GGGCTTACCA ATCAAATTAC	180

AATAATATGA	GATTTTAAAT	TTGTAATTTA	AAAAAATATT	AGTTTCTTAA	ATCATGTAGA	240
AAACAAAAAG	TAAAGTCACA	CATC	ATG	ATT	ACA	291
			Met	Ile	Thr	
			Met	Met	Leu	
					Ala	
					Leu	
					Ile	
					-10	
AGT	GTC	TGC	CTA	TTT	GCV	315
Ser	Val	Cys	Leu	Phe	Ala	
				Phe	Trp	
-5				1		

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 346 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 191..235  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.9  
seq LLTLVQCSDLCP/CS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

ACTGACATTG	RTTATCCARA	ATGACTAACA	CTCTTCTACA	TTAAATAGGG	AAAGAGCCTG	60				
ATTACAGTCC	CTCCAACAGT	CTACACAAAT	ACCCCCCAAC	ACACACRTAT	RATTGAGGGT	120				
GAAGTGTTTT	CTAGATCATT	GCCTAAGTCC	CCCATTTGCT	TTCAGAATAG	ACCCAGGCRR	180				
GTAAAAGGGA	ATG TGG CTC TTA ACA CTA GTT CAA TGT TCT GAC CTK TGT	229								
	Met Trp Leu Leu Thr Leu Val Gln Cys Ser Asp Leu Cys									
	-15 -10 -5									
CCT TCC TGC TCC CAA GCA TTG ACA CTT GTG TTA GTA TCT TTT TCT GAA	277									
Pro Ser Cys Ser Gln Ala Leu Thr Leu Val Leu Val Ser Phe Ser Glu										
	1 5 10									
GTC AGA GAC TTG GCA GAG ACC TCC CTA TCA TCT AAT CTG AAG AAC TCT	325									
Val Arg Asp Leu Ala Glu Thr Ser Leu Ser Ser Asn Leu Lys Asn Ser										
	15 20 25 30									
TTG TTT ATA GTT CTG AAG AGG	346									
Leu Phe Ile Val Leu Lys Arg										
	35									

## (2) INFORMATION FOR SEQ ID NO: 71:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 179 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 27..119
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9  
seq LLFACLTMLLVKT/CQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

ATTACCAGTG TACACTTGAG AGAGCT ATG AGA GTA CAT CTT TTC CCA TAC CTT	53
Met Arg Val His Leu Phe Pro Tyr Leu	
-30 -25	
TGT CAA CCT AGT GTA CTA TCA AAC TTT TTG TTA TTT GCT TGT CTT ACT	101
Cys Gln Pro Ser Val Leu Ser Asn Phe Leu Leu Phe Ala Cys Leu Thr	
-20 -15 -10	
ATG TTG TTG GTG AAA ACG TGT CAG GAG TCC CCA AAA TCA CCC CTA AGT	149
Met Leu Leu Val Lys Thr Cys Gln Glu Ser Pro Lys Ser Pro Leu Ser	
-5 1 5 10	
TTG ATG ATT TGC CAG ACT TAT AGA ATT GGG	179
Leu Met Ile Cys Gln Thr Tyr Arg Ile Gly	
15 20	

## (2) INFORMATION FOR SEQ ID NO: 72:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 261 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 121..165
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.7

seq PLCFLILPYPVLS/SH

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

```

ATTTYGGGTC TACGCATTCT GCTACTACTC CTGTTTTCCT CTACAACGTC CCATCTTTAT    60
CCTAGTGATG TGTCTACTTT ACAACATATA TACTCCATCT CCAGCCTCAG CAGCCCTGTC    120
ATG ATT CCC CTC TGT TTC CTC ATC TTA CCC TAT CCC GTG CTT TCC TCT    168
Met Ile Pro Leu Cys Phe Leu Ile Leu Pro Tyr Pro Val Leu Ser Ser
-15          -10          -5          1
CAT GAC CAT AAT TCC CTC GGT CTC TTA GCT GAC AAA GTA GCC AAT GAA    216
His Asp His Asn Ser Leu Gly Leu Leu Ala Asp Lys Val Ala Asn Glu
          5          10          15
ATT AAC CGA AGT AAT TGC CGG GTC TAT GCC CAT TCC CAC TCC GGG    261
Ile Asn Arg Ser Asn Cys Arg Val Tyr Ala His Ser His Ser Gly
      20          25          30

```

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 144 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 28..123
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.5  
seq CLLSXPSTRKSQA/CM

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

```

ACCTGCTCTA GCGGGCCGCG TAGACCA ATG GCG GGC TCC CGG CTC CCG CGG CAG    54
          Met Ala Gly Ser Arg Leu Pro Arg Gln
          -30          -25
CTC TTC CTC CAG GGC GTG SMG GCG TCT TCA TGT TTG CTT TCR MTT CCC    102
Leu Phe Leu Gln Gly Val Xaa Ala Ser Ser Cys Leu Leu Ser Xaa Pro
      -20          -15          -10
TCT ACA CGC AAA TCC CAG GCC TGT ATG GCC CCG AGG GCA TGG    144
Ser Thr Arg Lys Ser Gln Ala Cys Met Ala Pro Arg Ala Trp
      -5          1          5

```

## (2) INFORMATION FOR SEQ ID NO: 74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 250 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 146..229
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.4  
seq FVLHLLAQDLVCC/FY

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

```

ATGGCCTTTA TTTCTGTTCT TACTCCAACA TTTTCTCATC TTTTCTCCCA TCCTTTTACT    60
TAGAAAAATG TTAAATTAG AGGAAATAGT ACAATGAACA TGAGTATTCT CTTAATCTT    120
ATTCAATTAA TATATTTCTC TGTAT ATG TAT ATA TGC TTT TGT TTG GAA TCA    172
                        Met Tyr Ile Cys Phe Cys Leu Glu Ser
                        -25                               -20

TTT GAA ATT AAG TGT GGA TTT GTT CTC CAT CTT CTT GCT CAG GAT TTG    220
Phe Glu Ile Lys Cys Gly Phe Val Leu His Leu Leu Ala Gln Asp Leu
                        -15                               -10                               -5

GTG TGC TGC TTT TAT CTG AGG ACA TNN BAR                                250
Val Cys Cys Phe Tyr Leu Arg Thr Xaa Xaa
                        1                               5

```

## (2) INFORMATION FOR SEQ ID NO: 75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 231 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 127..186
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.4

seq LNAFTLLVWLSLS/KN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

```

ATTTTAATAT AATGCATTAA AATGTCAGGT AATACTGTAT ATTCTATATT GCATCACAAC      60
AGGAGATATA TCTGGATGAC CTACCATTAG TGATGCTAAG TTTTACATTG TATTGGAGCA      120
ACACCA ATG CAT TTC ATC CTC CAT AAC CTT AAT GCT TTT ACT CTC CTA      168
      Met His Phe Ile Leu His Asn Leu Asn Ala Phe Thr Leu Leu
      -20                -15                -10

GTT TGG CTA TCT CTT TCT AAA AAT ACA GTT CCC AGA CCA GCA GTA TTA      216
Val Trp Leu Ser Leu Ser Lys Asn Thr Val Pro Arg Pro Ala Val Leu
      -5                1                5                10

GCA TCG GCA GCC TGG                                          231
Ala Ser Ala Ala Trp
                        15

```

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 194 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (E) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 84..188
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.2  
seq IAPLFTLLPKSIP/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

```

ACACTCATTC TATACTGGGC ACCATGTCTG ATCATTTTTT TTTCTGATAT TTGGATCAGT      60
TCTATTAAAC TGATAACCCT GTG ATG TCT TTT TTT CCC TTC AAT AGA TCT TTA      113
      Met Ser Phe Phe Pro Phe Asn Arg Ser Leu
      -35                -30

AAT TCC AAT CCT CAC CCT AAT CTA CTC TTT CCC AAT ATA GCA CCG TTA      161
Asn Ser Asn Pro His Pro Asn Leu Leu Phe Pro Asn Ile Ala Pro Leu
-25                -20                -15                -10

TTC ACA CTG CTC CCA AAA TCT ATT CCA GCC CCG      194
Phe Thr Leu Leu Pro Lys Ser Ile Pro Ala Pro
      -5                1

```



## (2) INFORMATION FOR SEQ ID NO: 77:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 169 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 8..76
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.2  
seq LLDLHCFCSLAKT/KN

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

AAACAGA ATG GTT GTA TGG GTA CTT GAA GTA CGG TTT CTC CTG GAT TTG	49
Met Val Val Trp Val Leu Glu Val Arg Phe Leu Leu Asp Leu	
-20 -15 -10	
CAT TGC TTT TGC AGC CTT GCA AAG ACA AAA AAT GGT TTA AGT TGG GGA	97
His Cys Phe Cys Ser Leu Ala Lys Thr Lys Asn Gly Leu Ser Trp Gly	
-5 1 5	
CTG CCA CAA AAA GTG GCA CTC TGC ACA CCC TGT TCT GCA CCG GCT TTG	145
Leu Pro Gln Lys Val Ala Leu Cys Thr Pro Cys Ser Ala Pro Ala Leu	
10 15 20	
TTT TGG TTT GGT TTT CAC ATA CTG	169
Phe Trp Phe Gly Phe His Ile Leu	
25 30	

## (2) INFORMATION FOR SEQ ID NO: 78:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 351 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 223..294

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 5.2  
seq PGLCCPALGSAWS/KN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

```

ACTCCCTGAA CCCTAACTCA CCCTTTGCCC TCCCCACCCT TCAGCCCCTG CCCAGGTCTT    60
GGAGATCTCT GTGCTGTCTT TTGTGGAGCA GCTGCTATCT TGCAGTCAGA TCCTCTGTCTG    120
GGGAGGCCTT CAGCTTTTGT TCAACGACCC AGAGGGTGTG GGAGGGGCTC AGTTACTCTT    180
CTCCTCACCT GGCACCTAGA GAAAGCAAGT CTCAAGAGTC TC ATG GTA TGT GGT        234
                               Met Val Cys Gly
TGG TGG ACC CAG GGG CCT GTG CCC GGT CTG TGC TGT CCA GCT TTG GGC        282
Trp Trp Thr Gln Gly Pro Val Pro Gly Leu Cys Cys Pro Ala Leu Gly
-20                               -15                -10                -5
TCT GCC TGG AGC AAA AAC AAG AGC NTG CCT GTG CCG TGT TGC GGT CCT        330
Ser Ala Trp Ser Lys Asn Lys Ser Xaa Pro Val Pro Cys Cys Gly Pro
                               1                   5                   10
TAC ATG GTA GCG AAT CTC GGG                                            351
Tyr Met Val Ala Asn Leu Gly
                               15

```

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 362 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 231..308
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.2  
seq FECALVSASLTTA/GT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

```

AATTTTCAAA AGTGCTGTTA ACATAAACAG AGCAGTAAAT CTGGGGCACC ATGCTTTTTT    60
TTCAAAGTTG TTAAGAATTA TATCAGTCTG CAGGTGTCAC ACGCAGTTAC TCAGRATCAG    120
AWAGAAGGCT GATCGGGGGT TAGATCTCCA TCTATCTATC TTTTGTCAAC CAACCACGTC    180
CAGGCTGTTT ATTTAATACT TCCTCTTGCT AATGAAGGTA CTGGTTGGGG ATG GGT        236

```

Met Gly  
-25[illegible]

(2) INFORMATION FOR SEO ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 218 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Placenta

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 150..203  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5  
seq LCXXLLLCVLFVSH/FY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

AATY	TGAATT	GAAA	ATTAGC	TTCAT	GTTTGT	TAAGAT	GATC	ATAT	CACCTG	AGAG	AGTTCC	60			
CAAG	TCYACA	ATTG	CTCTAC	TAGTT	TACTAT	TCAGT	GTTTGT	TSAAAA	AATHT	TAAT	CTCAGT	120			
ACTG	TGAAGA	AGCT	GGAAAA	AGGG	GATATT	ATG	GGG	CTA	AAA	GCT	CTC	TGT	TTS	173	
						Met	Gly	Leu	Lys	Ala	Leu	Cys	Xaa		
									-15						
SGG	CTG	CTT	TGT	GTT	CTT	TTT	GTC	TCT	CAT	TTT	TAC	ACA	CCC	ACT	218
Xaa	Leu	Leu	Cys	Val	Leu	Phe	Val	Ser	His	Phe	Tyr	Thr	Pro	Thr	
-10					-5					1				5	

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 363 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: 82..309

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 5  
seq FPLLALLFEKCEQ/ST

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

```

AAGTGTGATG AAGATTGGCA CCCAGACACC ATTCGCTTTT CACCCAAGAT GATTTGATGT      60
CTTATAAAAC TCTGATGAAC C ATG ATG GCT ACA CAG ACA TTA AGT ATA GAC      111
                  Met Met Ala Thr Gln Thr Leu Ser Ile Asp
                  -75                               -70

AGC TAT CAA GAT GGG CAA CAG ATG CAA GTA GTA ACA GAG TTA AAG ACA      159
Ser Tyr Gln Asp Gly Gln Gln Met Gln Val Val Thr Glu Leu Lys Thr
-65                               -60                               -55

GAA CAA GAT CCA AAC TGC TCT GAA CCC GAT GCA GAA GGA GTG AGC CCT      207
Glu Gln Asp Pro Asn Cys Ser Glu Pro Asp Ala Glu Gly Val Ser Pro
-50                               -45                               -40                               -35

CCC CCT GTG GAG TCT CAG ACC CCG ATG GAT GTG GAC AAG CAG GCC ATT      255
Pro Pro Val Glu Ser Gln Thr Pro Met Asp Val Asp Lys Gln Ala Ile
-30                               -25                               -20

TAT AGG CAT CCA CTA TTT CCA TTA TTA GCT TTG TTG TTT GAA AAA TGT      303
Tyr Arg His Pro Leu Phe Pro Leu Ala Leu Leu Phe Glu Lys Cys
-15                               -10                               -5

GAA CAA TCT ACA CAG GGC TCT GAA GGC ACA ACT TCT GCC AGT TTT GAT      351
Glu Gln Ser Thr Gln Gly Ser Glu Gly Thr Thr Ser Ala Ser Phe Asp
      1              5              10

GTA GAC ATC GGG      363
Val Asp Ile Gly
15

```

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 258 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 163..225  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.9  
seq SVFLSGSVCLSFL/SE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

```

ATGAGGTTTT AGTTCTTTGG GCAGTAGCAA GTAGTGATAG GTGATGGTGG TCTGTGGTTC    60
ATGGACCAAAA AAATGTTTGA GAGGCATGGA GGATCTTTAG ACAAGATCAC CTTCTAGGTG    120
TATGTGGTAG CATTCTGTTC CACCGCTCTC TTCTCCTTCA GG ATG TCT CCC TCC    174
                               Met Ser Pro Ser
                               -20

CAG CTA ACC TGC TCG GTG TTC CTC AGT GGG AGC GTT TGC CTT AGC TTT    222
Gln Leu Thr Cys Ser Val Phe Leu Ser Gly Ser Val Cys Leu Ser Phe
   -15                      -10                      -5

CTC TCA GAG CAT CGT ACT TAC TTT TTC TGC CCA CTG    258
Leu Ser Glu His Arg Thr Tyr Phe Phe Cys Pro Leu
   1                      5                      10

```

## (2) INFORMATION FOR SEQ ID NO: 83:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 37..126  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.8  
seq FCSLLCLRTQLFP/HG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

```

AAAAACATGT CCTGGGTTGG CAGCTTGAGA TGATAA ATG CTT CAA GCC CTA GCC    54
                               Met Leu Gln Ala Leu Ala
                               -30                      -25

CCG GCA CAC CAC TTA TGC TCC CTG AAG AGG TCA TTC TGT TCT CTT CTG    102
Pro Ala His His Leu Cys Ser Leu Lys Arg Ser Phe Cys Ser Leu Leu

```

	-20	-15	-10	
TGC CTT CGC ACA CAG CTC TTC CCC CAC GGG				132
Cys Leu Arg Thr Gln Leu Phe Pro His Gly				
-5		1		

## (2) INFORMATION FOR SEQ ID NO: 84:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 487 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 383..424
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.8  
seq LFLKYLWRS LCRG/GI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

```

AGATATTTGG TGTGTCGTAG GCTTTGTTGC GAGTGGACTC TTGCATCATA TGTGGCTGTT    60
GGAGAATTGA TTCCAGCCT TTTGCGCTGG CTGAGACCTT GTATTCCAGT TCTAGGCATC    120
ACATCCAACCT CAACAGTGTC CATGGGAAAA AGAGCATCTC TTAAAGCAA GAAAACCTTT    180
TCCAGAAGCT TCCTGCAGAC CTCCCTTCAC ACCTCGTTGA CCAGGCTGAT GGTGTGTGCC    240
AGTGCCTAAG TCAAGTCATT GACAACAGGA AGAAGATGAG CCCTGGCTGA GTTAACTGA    300
CCAGGATGTG CATCCCCCCC ATGCTGGAAG TGAGGCCGTC TTCCCTGAGA GTCAGTATCC    360
GAACAGAAAG ATCACCCAAT GC ATG TTG TTT TTA AAG TAC TTA TGG AGA TCT    412
                        Met Leu Phe Leu Lys Tyr Leu Trp Arg Ser
                        -10                               -5

CTG TGC CGT GGT GGT ATC ATC CGT ATG AAC CAT CCA GGC TGT AGT CAG    460
Leu Cys Arg Gly Gly Ile Ile Arg Met Asn His Pro Gly Cys Ser Gln
                        1                               5                               10

AGA ATC AGA GAC TCG CTG TGT GAT CTC    487
Arg Ile Arg Asp Ser Leu Cys Asp Leu
                        15                               20

```

## (2) INFORMATION FOR SEQ ID NO: 85:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 296 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 12..122
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.8  
seq AILIRPLVSVSGS/GP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

ACTCGCTCAA G ATG GCG CTG CTC GCG ATG CAT TCT TGG CGC TGG GCG GCC	50
Met Ala Leu Leu Ala Met His Ser Trp Arg Trp Ala Ala	
-35 -30 -25	
GCG GCG GCT GCT TTC GAA AAG CGC CGG CAC TCC GCG ATT CTG ATC CGG	98
Ala Ala Ala Ala Phe Glu Lys Arg Arg His Ser Ala Ile Leu Ile Arg	
-20 -15 -10	
CCT TTA GTC TCT GTT AGC GGC TCA GGT CCG CAG TGG AGG CCA CAT CAA	146
Pro Leu Val Ser Val Ser Gly Ser Gly Pro Gln Trp Arg Pro His Gln	
-5 1 5	
CTC GGC GCC TTG GGA ACC GCT CGA GCC TAC CAG ATT CCA GAG TCA TTA	194
Leu Gly Ala Leu Gly Thr Ala Arg Ala Tyr Gln Ile Pro Glu Ser Leu	
10 15 20	
AAA AGT ATC ACA TGG CAG AGA TTG GGA AAA GGC AAT TCA GGA CAG TTC	242
Lys Ser Ile Thr Trp Gln Arg Leu Gly Lys Gly Asn Ser Gly Gln Phe	
25 30 35 40	
TTA GAT GCT GCA AAG GCT CTC CAG GTA TGG CCA CTG ATA GAA AAG AGG	290
Leu Asp Ala Ala Lys Ala Leu Gln Val Trp Pro Leu Ile Glu Lys Arg	
45 50 55	
ACT TGG	296
Thr Trp	

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 211 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Placenta

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 65..178  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.8  
seq LASLFGLDQXAAG/HG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

```
AACGACGTCG GCGGTGACAG GCCCGTGGGA CTYTGGGRAT ACCCAGCKTC CTCCCCGCAA      60
CCCG ATG AAA GCC ARC GCA ATG TTC GGT GCG GGG GAC GAG GAC GAC ACC      109
Met Lys Ala Xaa Ala Met Phe Gly Ala Gly Asp Glu Asp Asp Thr
           -35                      -30                      -25

GAT TTC CTC TCG CCG AGC GGC GGT GCC AGA TTG GCC TCA CTT TTT GGA      157
Asp Phe Leu Ser Pro Ser Gly Gly Ala Arg Leu Ala Ser Leu Phe Gly
           -20                      -15                      -10

CTG GAT CAG GSA GCT GCT GGC CAT GGA AAT GRA TTK TTC CAG TAC ACA      205
Leu Asp Gln Xaa Ala Ala Gly His Gly Asn Xaa Xaa Phe Gln Tyr Thr
           -5                      1                      5

GCC CCA
Ala Pro
10
```

## (2) INFORMATION FOR SEQ ID NO: 87:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 277 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 215..253  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.7  
seq MLWLLRSLTDVSS/MI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

```
AGCTCTTCTG AAGGTCAGGT CATCTAACTC CCTTGACCTC CTCCTTGAAG CTGTGAGCTA      60
CTTCTTGCG ATGCCCTCC ATGCAGGAGA GACTTGCACT TAGCTTCCTG CTTTCCCTT      120
```



CCTCCCCAAC TCCTGCCATG TCATTTTAC CCATTTTGT GACCTCAGCA TTGATTGGTT 180  
 GGAACCACCT AACGCCATGT AGTTGAACCA TTCA ATG CTC TGG CTT CTT AGG TCT 235  
 Met Leu Trp Leu Leu Arg Ser  
 -10  
 TTA ACA GAT GTG TCC TCA ATG ATC TTT TTC ACC ATT CCA GGG 277  
 Leu Thr Asp Val Ser Ser Met Ile Phe Phe Thr Ile Pro Gly  
 -5 1 5

## (2) INFORMATION FOR SEQ ID NO: 88:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 206 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 57..101
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.7  
seq IFHVLIHSSSFS/CE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

AAATTAAAA AAAAGTCAGT GCGTTTTGGG TATCATTATA GTTTAAATTT TCCTTA ATG 59  
 Met  
 -15  
 ACG ATT TTC CAT GTG CTT ATT GCC CAT TCA TCC AGC TTC TCT TGT GAA 107  
 Thr Ile Phe His Val Leu Ile Ala His Ser Ser Ser Phe Ser Cys Glu  
 -10 -5 1  
 GTT ATT GTT AAG GTC TTT TGT CCA TTT TTG GGT TGT TTA TCT TTT CAT 155  
 Val Ile Val Lys Val Phe Cys Pro Phe Leu Gly Cys Leu Ser Phe His  
 5 10 15  
 TAT CTT TAT TCA CTT TTG GAG TTC TTT ATT CTG AAT ACA AGT CCT TCG 203  
 Tyr Leu Tyr Ser Leu Leu Glu Phe Phe Ile Leu Asn Thr Ser Pro Ser  
 20 25 30  
 ATG 206  
 Met  
 35

## (2) INFORMATION FOR SEQ ID NO: 89:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 186 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 85..129
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.7  
seq WQLXGFCGSYSA/AQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

```
AACTTTCGCT GAGGWGCCGA GCGACGGGG CTTGGGTTW YYTCCATTAT ACTCATGTAA    60
TGTTCTTGGT TTGATTTATT CAGT ATG CAC TGG CAG CTT TTG BBA GGC TTC    111
               Met His Trp Gln Leu Leu Xaa Gly Phe
               -15                               -10

TGT GGG AGC TAC TCG GCT GCC CAA GCC GAG GCA CAA ACC CTG CCA GGT    159
Cys Gly Ser Tyr Ser Ala Ala Gln Ala Glu Ala Gln Thr Leu Pro Gly
   -5              1              5              10

CTC CAT AGT AAA TAC AAC ACG CAC GGG    186
Leu His Ser Lys Tyr Asn Thr His Gly
              15
```

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 308 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 114..224
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.6  
seq FVLLFFFSXLXY/FM

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

```

ACAGCTCTCT AAAGCCGAGA TAAACCTCTA CTTGCAAAGG GAATTTGAAA AAGATGAGAA    60
GCCACGTGAC AAGTCATATC AGGATGCAGT TTTAGAAGAT AWTTTTRAGR AGA ATG      116
                                     Met
ACC ATG ATG GTG ATG GCT TCA TTT CTC CCA AGG AAT ACA ATG TAT ACC      164
Thr Met Met Val Met Ala Ser Phe Leu Pro Arg Asn Thr Met Tyr Thr
   -35                               -30                       -25

AAC ACG ATG AAC TAT AGC ATA TTT GTA TTT CTA CTT TTT TTT TTT AGC      212
Asn Thr Met Asn Tyr Ser Ile Phe Val Phe Leu Leu Phe Phe Phe Ser
   -20                               -15                       -10                       -5

WAT TTA SKG TAC TTT ATG TAT AAA ACM AGT CAC TTT TCT CCA AGT TGW      260
Xaa Leu Xaa Tyr Phe Met Tyr Lys Thr Ser His Phe Ser Pro Ser Xaa
                               1                               5                               10

ATT TGC TAT TTT TCC CCT ATG AGW AGH WAT TTK GAT CTC CCC AAT GGG      308
Ile Cys Tyr Phe Ser Pro Met Xaa Xaa Xaa Xaa Asp Leu Pro Asn Gly
   15                               20                               25

```

## (2) INFORMATION FOR SEQ ID NO: 91:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 438 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 139..321
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.5  
seq LVTRLALCQSPRA/GQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

```

AAGGGAGNNA GGAAGCTGCG CTGCATTCTG CGGGACGAAC CCTGCTCCGC GCGAGAATTT    60
TTTTGATTCC TTCTTATTTG GAGAAATCTC CAGCTGCTCT GATGATAGCC TAAGAAGACT    120
GCATGCTGCT TCCTCTCG ATG CCA AGC CAG ACC CTC TCA CAA CCT CGG ATC      171
               Met Pro Ser Gln Thr Leu Ser Gln Pro Arg Ile
               -60                               -55

TCA GTC CTT CAT GGA GAC CTG GTC CCA GCA GGA ATG GCA GTG CAG GAA      219
Ser Val Leu His Gly Asp Leu Val Pro Ala Gly Met Ala Val Gln Glu
   -50                               -45                       -40                       -35

ATT GGC GCC CAG ATG GTT CTT CCA TGT GAA GTT GTC TCG GGC TCT GGG      267
Ile Gly Ala Gln Met Val Leu Pro Cys Glu Val Val Ser Gly Ser Gly

```

	-30	-25	-20	
CTG ACG AGA GAA CAC CTG GTA ACC AGG TTA GCC CTC TGT CAG TCA CCC				315
Leu Thr Arg Glu His Leu Val Thr Arg Leu Ala Leu Cys Gln Ser Pro				
	-15	-10	-5	
AGG GCA GGG CAG CAT GGT GCG GAT TCA GAG GAG GAA GCT TTT GGC ATC				363
Arg Ala Gly Gln His Gly Ala Asp Ser Glu Glu Glu Ala Phe Gly Ile				
	1	5	10	
TTG CCT GTG CGT CAC AGC CAC CGT CTT TCT GCT TGT CAC ACT CCA GGT				411
Leu Pro Val Arg His Ser His Arg Leu Ser Ala Cys His Thr Pro Gly				
	15	20	25	30
GAA CTG AGG TTT AGT GAA TGG ACG TGT				438
Glu Leu Arg Phe Ser Glu Trp Thr Cys				
	35			

## (2) INFORMATION FOR SEQ ID NO: 92:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 278 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 147..260
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.5  
seq LLMITVTVGPGAS/GV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

AAGACACGGA CAGACGGACG CGCAGACCTT CGGAAGGCAC TGCCTAGGCA GCCTCCCCGG	60
AGCCCACGAG GCTCCCCAGC ACCGTTCACT GGTGGGAGGC TGAGCCGGTG GAAAAGACAC	120
CGGGAAGAGA CTCAGAGGCG ACCATA ATG TCG TTA CGT GTA CAC ACT CTG CCC	173
Met Ser Leu Arg Val His Thr Leu Pro	
	-35 -30
ACC CTG CTT GGA GCC GTC GTC AGA CCG GGC TGC AGG GAG CTG CTG TGT	221
Thr Leu Leu Gly Ala Val Val Arg Pro Gly Cys Arg Glu Leu Leu Cys	
	-25 -20 -15
TTG CTG ATG ATC ACA GTG ACT GTG GGC CCT GGT GCC TCT GGG GTG TGC	269
Leu Leu Met Ile Thr Val Thr Val Gly Pro Gly Ala Ser Gly Val Cys	
	-10 -5 1
CCC TCT GGG	278

Pro Ser Gly  
5

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 335 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 276..329
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.5  
seq SLLLLGRWLTLS/SG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

```
AATTTTGGHA GTAGAGTCTG TAGTGGTACT AATTAAAATA GTCATGTAGA TAAACTGTTT    60
CTTACTGTTT CAGGACAAGA TGAGTCTATG CCAGATGTAA GTCACCACAC TCATTTACTG    120
ATTTCTTCCT TGAGGATTTT TTGTTGTAA AAAATTTTGT TGGTTGATCT GCATTCTGTC    180
TTAAACTCTT TATTTTATTG TGGATTATTA ACAGTTGGTC CACGGGCCAC ACTTTAAGCA    240
ACGTCAATCT CTAGTATCCA TCTTAGAGGC AAACC ATG ATC TAT TTA ACC AGT      293
                               Met Ile Tyr Leu Thr Ser
                               -15

CTC TTA CTA TTG GGT AGA TGG TTA ACC CTC ACA TCA TCT GGG              335
Leu Leu Leu Leu Gly Arg Trp Leu Thr Leu Thr Ser Ser Gly
-10                               -5                               1
```

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 363 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 292..357
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.4  
seq GFLLCPLVCGLRR/WT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

```

AACCTCTCTT CCTTTCCTGA GTCGTTGGCG CAGTGA CTCG CCACCTCCTC CCTCCTCGTC    60
CCCCACCGG AGGAGTTTTG CGGTCTGTAG AGAGCTATGC AGGGTGAGGG CCCCTAGACT    120
TCGGCTTTTCG CCGCTGTTGG TG GTAGGCTG GAGTTGGGGG TCCCTGGATA CGGTTTTGGC    180
TTTACACCCC CTTTCCAGCA AGCTTCCCGT GGAATCTGT TCCTTTTCAG GACAGCTGTG    240
ATCCACGTGA GTAAACTGG GCGTCCGTC TTGTGCTTTT TCCTCAGGTT C ATG AAC    297
                                     Met Asn

TGG AAT GTA AGA GGC ACC AGA GGA TTC CTG CTC TGT CCC CTG GTT TGC    345
Trp Asn Val Arg Gly Thr Arg Gly Phe Leu Leu Cys Pro Leu Val Cys
-20                      -15                      -10                      -5

GGC TTG CGA CGT TGG ACA                                363
Gly Leu Arg Arg Trp Thr
1

```

## (2) INFORMATION FOR SEQ ID NO: 95:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 261 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 130..255
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.4  
seq LVCLTFITATTHE/QP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

```

AATTGGGTGC ATACTTTTTC TTGGTCAGCA TCTCAGTTTC TATTTCTAAA GTCAAAAGAA    60
GTGCCTCGAT TCCTCCCAAG ACTTGACAAG CCCAGCAGGG TTGGTGGCTT CTTTGCTGTG    120
TTGCTGAAA ATG GAA CAG GCT GCC CTG GAG GTG GTG AGC CCC CTG CCC CGG    171
Met Glu Gln Ala Ala Leu Glu Val Val Ser Pro Leu Pro Arg

```

	-40		-35		-30	
AGA TGT TCA GTG AGA TCA CCT GTG ACT ACA TGC TGT GCT AAG GAC CTT						219
Arg Cys Ser Val Arg Ser Pro Val Thr Thr Cys Cys Ala Lys Asp Leu						
	-25		-20		-15	
GTG TGC CTC ACC TTC ATC ACT GCA ACA ACC CAT GAG CAG CCG						261
Val Cys Leu Thr Phe Ile Thr Ala Thr Thr His Glu Gln Pro						
	-10		-5		1	

## (2) INFORMATION FOR SEQ ID NO: 96:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 323 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 24..275
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.4  
seq GLVQLHATXLALG/KV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

AATACTCTCA TCTGTAAAAT GGG ATG ATA ATC CCA CTA CCC AGC TTA GTG GGC	53
Met Ile Ile Pro Leu Pro Ser Leu Val Gly	
	-80 -75
TGT TGG GAA GGA GGG AAT GGG AAA GGA CTT ATG GTG TCT GAC ACT ACA	101
Cys Trp Glu Gly Gly Asn Gly Lys Gly Leu Met Val Ser Asp Thr Thr	
	-70 -65 -60
TGC TGG ACA CTC GCT TCC TCC AAT GTC CCA TCT CCR TCC CCT GCG CCC	149
Cys Trp Thr Leu Ala Ser Ser Asn Val Pro Ser Pro Ser Pro Ala Pro	
	-55 -50 -45
ACC CTG GGG AGA GGN GCC CCC TCC CAT ACT CCC CAG AAG AAG CCC ACC	197
Thr Leu Gly Arg Gly Ala Pro Ser His Thr Pro Gln Lys Lys Pro Thr	
	-40 -35 -30
ATA CCT GGT GCC CGC CAC CGC CCC ATC ATT CTT CCC AAG GGG CTC GTC	245
Ile Pro Gly Ala Arg His Arg Pro Ile Ile Leu Pro Lys Gly Leu Val	
	-25 -20 -15
CAG CTC CAC GCC ACA YCA CTC GCC CTT GGC AAA GTC TGT CTC CCC CAC	293
Gln Leu His Ala Thr Xaa Leu Ala Leu Gly Lys Val Cys Leu Pro His	
	-10 -5 1 5
GTA CCG CAC CAC GCY AGT CTT CGT CCC GCA	323

Val Pro His His Ala Ser Leu Arg Pro Ala  
                   10                                  15

## (2) INFORMATION FOR SEQ ID NO: 97:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 235 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 65..190
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.3  
seq IQTVHIALPGSLG/HP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

ATATGGTAAT TAMVAATATG TTTATGTACC CTGAATCATG TAAATATTTG AGCTTTCTCT	60
AAAA ATG AGT ATG AGA CTA TCT GGA GAA AGA ATT TAT CTC CTG TTA GAG	109
Met Ser Met Arg Leu Ser Gly Glu Arg Ile Tyr Leu Leu Leu Glu	
-40                                  -35                                  -30	
GTT TGG CTG CCT TRW CTC AAT TTT GAG TCA GTT CTT CAT TTT ATC CAA	157
Val Trp Leu Pro Xaa Leu Asn Phe Glu Ser Val Leu His Phe Ile Gln	
-25                                  -20                                  -15	
ACT GTC CAC ATT GCC CTC CCT GGA AGT CTG GGC CAC CCA ATG GGC CCC	205
Thr Val His Ile Ala Leu Pro Gly Ser Leu Gly His Pro Met Gly Pro	
-10                                  -5                                  1                                  5	
TGT GCC TGC CGC CCC TCT TTA GCC CAC CCG	235
Cys Ala Cys Arg Pro Ser Leu Ala His Pro	
10                                  15	

## (2) INFORMATION FOR SEQ ID NO: 98:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 197 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens



(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 144..191
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.3  
seq LLLFCFMPVVINP/DR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

```

ATCATTGTC AGAGAAGAGA GAAGGGTATG AGTTGCCAGG TAGTGGATGC AGATGGAAGC   60
GCTGGTGGGC CCATTGGTTG ATCATTGGTT GGGACCATCT TACACAGAAA GTTCATCCTA   120
TTGCCCTTTC CCACTGTGTT AAT ATG GGA ACA CTC CTT CTC TTC TGT TTT ATG   173
               Met Gly Thr Leu Leu Leu Phe Cys Phe Met
               -15                               -10

CCT GTA GTG ATT AAC CCG GAC AGA                               197
Pro Val Val Ile Asn Pro Asp Arg
   -5                               1

```

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 253 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 86..151
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.1  
seq GIYLQLFFLSIVS/QP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

```

AGCAGTGCVG AAGAGAAAAG CAGAGATAAG CAAGGCTCCG CAGCAGCCTC CTTTCTAACT   60
TGACCCTCGC CAGACCCTGG CCAGC ATG GTT GTC CTG AAT CCA ATG ACT TTG   112
               Met Val Val Leu Asn Pro Met Thr Leu
               -20                               -15

GGA ATT TAT CTT CAG CTT TTC TTC CTC TCT ATC GTG TCT CAG CCG ACT   160
Gly Ile Tyr Leu Gln Leu Phe Phe Leu Ser Ile Val Ser Gln Pro Thr
   -10                               -5                               1

TTC ATC AAC AGC GTT CTT CCA ATC TCA GCA GCC CTT CCC AGC CTG GAT   208

```

Phe Ile Asn Ser Val Leu Pro Ile Ser Ala Ala Leu Pro Ser Leu Asp  
           5                          10                          15  
 CAG AAG AAG CGT GGT GGC CAC AAA GCA TGC TGC CTG CTG ACG CCG 253  
 Gln Lys Lys Arg Gly Gly His Lys Ala Cys Cys Leu Leu Thr Pro  
      20                          25                          30

## (2) INFORMATION FOR SEQ ID NO: 100:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 358 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 212..319
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.1  
seq HWLFLASLSGIKT/YQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

ATCCCCAWNS CACTCTCTCA CAGAGACTGT TCTTTTCCTT CTGAGACCCT ACTCCAGCTT 60  
 GTAGTTCTAA ATCTGTGATT ATGCACTGTC TGTCTTCCTC TTGAGGTCAG GGGCCATTTC 120  
 TTTTGTTCTC TGCTATGCTC AGGACCCAGA TCAAAGGAGC TCAGTAACTA TTTACAGGCG 180  
 TACATCATAT GTGGAGGACA CTTATGCTGT G ATG GCC CCA CAC ACA GCT TCC 232  
   Met Ala Pro His Thr Ala Ser  
   -35                          -30  
 TTT GGG GTC TGT CCC CTG CTC TCC GTT ACC CGC GTG GTA GCC ACT GAG 280  
 Phe Gly Val Cys Pro Leu Leu Ser Val Thr Arg Val Val Ala Thr Glu  
                           -25                          -20                          -15  
 CAC TGG CTC TTC CTG GCT TCA CTC TCT GGC ATC AAA ACT TAT CAG TCC 328  
 His Trp Leu Phe Leu Ala Ser Leu Ser Gly Ile Lys Thr Tyr Gln Ser  
                           -10                          -5                          1  
 TAC ATC TCA GTC TTT TGC AAG GTG ACT GGG 358  
 Tyr Ile Ser Val Phe Cys Lys Val Thr Gly  
      5                          10

## (2) INFORMATION FOR SEQ ID NO: 101:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 181 base pairs

- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 113..172
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.1  
seq SLPCLSFCTLCLV/TP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

```
ATCCCATCTT AGCTGCCATC CCATCCTGTA TG TAGATACC CTGTT CATCT CTCTCAAGAT   60
CTTTATTCTA GACCTTCTTT GTCTTTCTTG GGCTCCAACA CCTCACACCA AG ATG TCT   118
                                         Met Ser
                                         -20

TAC AAG TGG ATG CCC TCC TTA CCC TGC TTG AGT TTC TGT ACC CTG TGC   166
Tyr Lys Trp Met Pro Ser Leu Pro Cys Leu Ser Phe Cys Thr Leu Cys
      -15                -10                -5

TTG GTC ACC CCC GGG   181
Leu Val Thr Pro Gly
      1
```

(2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 117..182
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.1  
seq LAGFLLVLYVCLP/HA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

```
AACTTTGTTG AAAGTGAGAA GACTGGCTTC CGTTACTCTC AATATGTTTA TTTCTGTGAT   60
```

CACCCCTTTTG TAATTAATCT CTGGTGTCCA CCGCCCACTG TTACCCAAGT GAATGC ATG 119  
Met

CCT CTC CCC ACC TGG GCT CCG ACC CTG GCA GGG TTC CTG CTT GTC TTG 167  
Pro Leu Pro Thr Trp Ala Pro Thr Leu Ala Gly Phe Leu Leu Val Leu  
-20 -15 -10

TAT GTC TGT CTC CCT CAC GCC GGG 191  
Tyr Val Cys Leu Pro His Ala Gly  
-5 1

## (2) INFORMATION FOR SEQ ID NO: 103:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 129 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 31..81
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.1  
seq LLDWIGLKALIRG/HD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

ACACTATGAT TTTATAAAAC AATTTTTTCT ATG AAC CTT TAC TTA CTT GAC TGG 54  
Met Asn Leu Tyr Leu Leu Asp Trp  
-15 -10

ATT GGA CTA AAA GCA CTG ATC AGA GGC CAC GAC ATA AAA ATT CAG TCC 102  
Ile Gly Leu Lys Ala Leu Ile Arg Gly His Asp Ile Lys Ile Gln Ser  
-5 1 5

CTT TGT CCT TCC CCG TGC CTC CCA AGG 129  
Leu Cys Pro Ser Pro Cys Leu Pro Arg  
10 15

## (2) INFORMATION FOR SEQ ID NO: 104:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 198 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 28..189  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.1  
seq LLLFCFMPVVINP/DX

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

```
ATCATTGGC AGAGAAGAGA GRWGGGT ATG AGT TGC CAK GTA SWK GAT GCA RSS      54
                               Met Ser Cys Xaa Val Xaa Asp Ala Xaa
                               -50

ARG CGC TGG TGG GCC CAT TSG TTG ATC ATT GGW TGG GRC CAT CTT ACA      102
Xaa Arg Trp Trp Ala His Xaa Leu Ile Ile Gly Trp Xaa His Leu Thr
-45                      -40                      -35                      -30

CAG AAA GTT CAT CCT ATT GCC CTT TCC CAC TGT GTT AAT ATG GGA ACA      150
Gln Lys Val His Pro Ile Ala Leu Ser His Cys Val Asn Met Gly Thr
                      -25                      -20                      -15

CTC CTT CTC TTC TGT TTT ATG CCT GTA GTG ATT AAC CCG GAC ARM GGG      198
Leu Leu Leu Phe Cys Phe Met Pro Val Val Ile Asn Pro Asp Xaa Gly
-10                      -5                      1
```

## (2) INFORMATION FOR SEQ ID NO: 105:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 148 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 59..124  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4  
seq ILAFQTFLNLRA/HL

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

```
AGATTACAAA AAAAAAATTC TCAAGTGAAT AAAGTCCATT GATAGATATT TTGTTTAA      53

ATG GTA CCA AAT CTG TGT GGA AGG CAA ATT TTG GCT TTC CAG ACA TTC      106
Met Val Pro Asn Leu Cys Gly Arg Gln Ile Leu Ala Phe Gln Thr Phe
```

-20	-15	-10	
TTG CTG AAC TTG AGA GCT CAT CTT TTT CAA CTG GCC TCC CGG			148
Leu Leu Asn Leu Arg Ala His Leu Phe Gln Leu Ala Ser Arg			
-5	1	5	

## (2) INFORMATION FOR SEQ ID NO: 106:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 82..123
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4  
seq FSLIIFFFPSPSSP/XA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

ACAACCTKHN TCTCMCTGT ATCTTTGTCA TGTCAGTCAC TGATATCAGC ATCTGCCCAG	60
TTGCTCAGGC CAAAACCTTA G ATG TTT TCC TTG ATT ATT TTT TTT TTC CCT	111
Met Phe Ser Leu Ile Ile Phe Phe Phe Pro	
-10	-5
CCC TCA TCC CCW AMR GCC AAT CCA TTT CCA TCC TAT CTT CAA AAT ATA	159
Pro Ser Ser Pro Xaa Ala Asn Pro Phe Pro Ser Tyr Leu Gln Asn Ile	
1	5
TTA TAC CTG AAA TTT GTC CAT MTC TCC CAT CTA TAC TGM MAC CCT CCG	207
Leu Tyr Leu Lys Phe Val His Xaa Ser His Leu Tyr Xaa Xaa Pro Pro	
15	20
	25
TCA GAG TGT GTT CAT ATT TCT AGT GGA TTA CCA	240
Ser Glu Cys Val His Ile Ser Ser Gly Leu Pro	
30	35

## (2) INFORMATION FOR SEQ ID NO: 107:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 331 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 242..310
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4  
seq LHCLLIVFILVEF/CK

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

```
ACATCCAATA ATCATTTATA AAGATCTCTA ACAGGCCAGT CAGTATACAG AGTACCAGAT   60
TAAAAATAAA TGTAGCACCA GTTTTTCAGA AATTATTATG TGTCTATAAT TAGGGTAATT  120
ACATTTAGAA GATCTTTTTG ATGATCTCCT TAAAGTCAGC AACTGTCTTT TTCATCTTTG  180
TTTACCTAGT ACCTGGAATG GAGATAGGCG TTTAGCACTT AAATGTTTAC TGMATATTCT  240
T ATG AGT GCC TTT TAT CTT TCC TAC TCC TTG TTG CAT TGC TTA CTT ATT  289
  Met Ser Ala Phe Tyr Leu Ser Tyr Ser Leu Leu His Cys Leu Leu Ile
    -20                -15                -10

GTT TTT ATT TTA GTT GAG TTT TGT AAG AAA TTG ACT TAC TTT           331
Val Phe Ile Leu Val Glu Phe Cys Lys Lys Leu Thr Tyr Phe
   -5                1                5
```

## (2) INFORMATION FOR SEQ ID NO: 108:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 315 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 160..249
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4  
seq SAGVVLTMDGASA/EQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

```
AGCAAACCTCT TAAGGTCACT TTCTGAAGGC GGCCTCATCA CAGTCGGAGG TATCATGATA   60
TTAGCTGGTT TGACATCAAG TCATTTGTGA GTCATCAGAT CTTCTCCTGA AAATGGGAGA  120
```

```

CACAGTAGGG CCCCTCCCAG GAGCTCTTGG CTGTTGCTG ATG GCA GAA GCC AAG      174
                      Met Ala Glu Ala Lys
                      -30

CTT GTC CAA GGT TCA CTT GTA GCC CCT CAG CGT CAS TCA GCT GGT GTC      222
Leu Val Gln Gly Ser Leu Val Ala Pro Gln Arg Xaa Ser Ala Gly Val
-25                      -20                      -15                      -10

GTC CTG ACC ATG GAC GGC GCG TCG GCC GAG CAA GAT GGC CTC CAG GAG      270
Val Leu Thr Met Asp Gly Ala Ser Ala Glu Gln Asp Gly Leu Gln Glu
                      -5                      1                      5

GAC AGA TCC CAC AGT GGC CCC TCG TCT CTC CCC GAG GCC CAC CGG      315
Asp Arg Ser His Ser Gly Pro Ser Ser Leu Pro Glu Ala His Arg
          10                      15                      20

```

## (2) INFORMATION FOR SEQ ID NO: 109:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 222 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 1..177
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4  
seq VLLTISTNASVLG/DG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

```

ATG AAA GGA GTA GGG CCT GAG CAG CTG AAT GAT GGA GCG CCA TCA AAT      48
Met Lys Gly Val Gly Pro Glu Gln Leu Asn Asp Gly Ala Pro Ser Asn
                      -55                      -50                      -45

GAG ATT GAA ATG ACT CCA TGT TTT TTC AGT GAG TTC CTT CTA TTG GAC      96
Glu Ile Glu Met Thr Pro Cys Phe Phe Ser Glu Phe Leu Leu Leu Asp
                      -40                      -35                      -30

GTT GGT GTT GTT AAT ATA GTA GTT ATT AAA ATG TCT TAT AAT GTC CTG      144
Val Gly Val Val Asn Ile Val Val Ile Lys Met Ser Tyr Asn Val Leu
                      -25                      -20                      -15

TTA ACG ATT AGT ACT AAT GCC TCT GTA CTT GGT GAT GGT GCT CAT AGA      192
Leu Thr Ile Ser Thr Asn Ala Ser Val Leu Gly Asp Gly Ala His Arg
                      -10                      -5                      1                      5

GTT ACT ACT AGA ATA AGA AGG CCA GGG GGG      222
Val Thr Thr Arg Ile Arg Arg Pro Gly Gly
                      10                      15

```



## (2) INFORMATION FOR SEQ ID NO: 110:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 464 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 255..389
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq QLFWVTASTFCRS/DI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

```

AAGATTAAGG AAAGCTGCCC TAGACAATGC CATGCCCTCA TTTTTCAGTT TCAGTCATTA    60
TTTTTCTCAG CAGTGTCTCTT TTGGCTTCAC TGTGTCTTCT AACATCATCA GCCAAATTGT    120
TTCTTTCTTT TGTAATCTGT AGTTTCAAAA TAATAGGAGT TGTTTTGCTT TCAGAAAAAG    180
ACAATTATAA TGTTGATTTG GTTCTTTTAA AAAACTAAAC ACACCCTCTG GAGATTCTAA    240
TTTACTCCGT ATTC ATG CTG AGG AAA CTA AGT GCC AGC AAT GAA AAC TTG        290
      Met Leu Arg Lys Leu Ser Ala Ser Asn Glu Asn Leu
      -45                      -40                      -35

TGT CTA CTC TCA AAC CCC TCT CAC AAT GAG GTC TAT TTG ATC AGA TGC        338
Cys Leu Leu Ser Asn Pro Ser His Asn Glu Val Tyr Leu Ile Arg Cys
      -30                      -25                      -20

TGT GAA TCC CAT CAG CTT TTC TGG GTA ACT GCC AGC ACA TTT TGC CGT        386
Cys Glu Ser His Gln Leu Phe Trp Val Thr Ala Ser Thr Phe Cys Arg
      -15                      -10                      -5

AGC GAT ATA GCT ACT ATG GCT AGT CTC CTG CCA TCT GTT CTT CTA ATG        434
Ser Asp Ile Ala Thr Met Ala Ser Leu Leu Pro Ser Val Leu Leu Met
      1                      5                      10                      15

CAA CTG TTC TCA ACA TTT TTT TTG AAT CTC                                464
Gln Leu Phe Ser Thr Phe Phe Leu Asn Leu
      20                      25

```

## (2) INFORMATION FOR SEQ ID NO: 111:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 268 base pairs

(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 215..259  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.8  
seq PLILLPLNPFVLQ/VA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

```
ATGATGGGGT GTTAGTAAAA GTGATACTAG TTCTTGATGA TGTTATCTC TTTGCCTTGT   60
ATTCAGGAGT TAAAAATTAA TGAGGGCCTT TTCTGAACTG TAAAGATATG GTATTTAGCA  120
GTTCTTAATA TACTGAGGGT TTCTCATTCT CTCTTTTGA TTATGTTGTA TTTGGCACAT  180
GGTTATTTGG GGACTACTTA CATTCTTAAA TTGA ATG TAT CCT TTG ATT CTC CTT   235
                        Met Tyr Pro Leu Ile Leu Leu
                        -15                      -10

CCT CTT AAC CCA TTT GTG CTG CAG GTT GCT GGG   268
Pro Leu Asn Pro Phe Val Leu Gln Val Ala Gly
      -5                      1
```

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 440 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 240..308  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.8  
seq RGFVAVGLGQISA/SP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

```
AGCAYSAGCA MAACTGACAA SYCGCSGCAG GAAGWYGTTG CCATCTCCAC TCGTTGGATT   60
```

```

TCTGTTGCTC TGGGTTGCAG CCTCTASGGG TGAGATCGTG CTGACTCAGT CTCCGGACTT 120
TCAGCCTGTG ACTCCTGGCG ASMCGCTCAC TATCACCTGC CGGGCCAGTC AAGACATAAG 180
TYACAAATTA CATTGGTACC AGCAGAAGCC TGGTCACTCT CCAAGGCTCC TCGTCAAAT 239
ATG CTT CTC AGA CCC TCT CCG GGG TCC CCT CGA GGT TTC GTG GCA GTG 287
Met Leu Leu Arg Pro Ser Pro Gly Ser Pro Arg Gly Phe Val Ala Val
      -20                      -15                      -10

GGA CTG GGA CAG ATT TCA GCC TCG CCA TCG ATG GCC TGC AAA CTG ACG 335
Gly Leu Gly Gln Ile Ser Ala Ser Pro Ser Met Ala Cys Lys Leu Thr
      -5                      1                      5

ATT TTG CAA CAT ACT TCT GTC TTC AGA GTA GTA GTC TTC CGT ACA CCT 383
Ile Leu Gln His Thr Ser Val Phe Arg Val Val Val Phe Arg Thr Pro
    10                      15                      20                      25

TTG GTC AGG GGA CCA CTC TCC AGG TCA AAC GAG CTG TGG CTG CAC CAT 431
Leu Val Arg Gly Pro Leu Ser Arg Ser Asn Glu Leu Trp Leu His His
      30                      35                      40

CTG TCT TCA 440
Leu Ser Ser

```

## (2) INFORMATION FOR SEQ ID NO: 113:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 167 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 93..146
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq ATACGPAAHQCSA/VP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

```

AATACTAATT CCTACCAGTC CTCACGAGGA GGGTGACACA CACCCTGTTC CCACTTGAGA 60
GAGGGCATCT GCAGGGGTGG AGGGGAGACC CC ATG GCT CGG CCT GGT GCC ACA 113
Met Ala Arg Pro Gly Ala Thr
                        -15

GCC TGC GGG CCT GCC GCC CAC CAG TGC TCT GCG GTC CCA CTG TGG TCC 161
Ala Cys Gly Pro Ala Ala His Gln Cys Ser Ala Val Pro Leu Trp Ser
    -10                      -5                      1                      5

```

CCT GGG  
Pro Gly

167

## (2) INFORMATION FOR SEQ ID NO: 114:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 259 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 191..247
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq LSLCIXXLEHLFT/WP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```
ACACTGAATC TTCTAGAGCC ACGGTAGCCA ACCTTTTAAA ATCAGTTGAG GGCACCTTAAA   60
AAAATACATC AGTATTTTTTA CACCTGATTT TTACACCTGG GCCCCAGCCT TGTGCAACAG   120
AACCTAGGGG TGGAGTCTAA GCATGGACAG TTTTAAAGC CCCAGGCAGC CAGGGCTGAG   180
GACCTTGGCG ATG GAG CCT GTW AGT TCG CTT TCC TTG TGT ATA TKG WCT       229
      Met Glu Pro Val Ser Leu Ser Leu Cys Ile Xaa Xaa
                -15                               -10

CTT GAG CAT CTC TTC ACA TGG CCC AAA GGG                               259
Leu Glu His Leu Phe Thr Trp Pro Lys Gly
      -5                               1
```

## (2) INFORMATION FOR SEQ ID NO: 115:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 203 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 120..176  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 3.8  
 seq WCSAAAWRSPLSA/AT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

```

ACTCGCCGGC GCCAGGCAGT GGAAGTCAG GTTCTTCGC CACCTCCCAG CCAGGACTCT   60
GCCACCCTCC CTCAGGATGC CTGAGGGCCC CGAGCTGCAC CTGGCCAGCC AGTKTGTGA   119
ATG AGG CCT GCA GGG CGC TGG TGT TCG GCG GCT GCG TGG AGA AGT CCT   167
Met Arg Pro Ala Gly Arg Trp Cys Ser Ala Ala Ala Trp Arg Ser Pro
          -15                      -10                      -5

CTG TCA GCC GCA ACC CTG AAG TGC CCT TTG AGA GGG   203
Leu Ser Ala Ala Thr Leu Lys Cys Pro Leu Arg Gly
          1                      5
  
```

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 66 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 10..57  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 3.7  
 seq CAYVLFFFNGCLY/RR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

```

ACTTGGGAA ATG TGG TTG TGT GCG TAT GTA TTA TTT TTT TTT AAT GGA TGT   51
Met Trp Leu Cys Ala Tyr Val Leu Phe Phe Phe Asn Gly Cys
          -15                      -10                      -5

TTA TAT AGG AGA AAG   66
Leu Tyr Arg Arg Lys
          1
  
```

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 289 base pairs

- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 152..196
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.6  
seq LLHRAVVLRLQQA/CR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

```

AAAAAATTGC AGTGCTGAAG ACGTGGACC CGCAAAGGC TGTCCCTCCC AACCTGGGA    60
TTCTGGGCTC ACTGAGTTCA CCTGCGAGTC AGCCCTACCT GCACTGCTCT GGTCTAGTAC    120
AAACAGGCTG CTGGCATTGA GGTCTGCTAC A ATG CTG CTG CTG CAC AGA GCT    172
                               Met Leu Leu Leu His Arg Ala
                               -15                      -10

GTG GTC CTC AGG CTC CAA CAG GCC TGC AGA CCG ACC TCT CTT CCA GAC    220
Val Val Leu Arg Leu Gln Gln Ala Cys Arg Pro Thr Ser Leu Pro Asp
          -5                      1                      5

TCA AGT CAA TCC CCT CAA GGA TCT GCA TTC AGG CCT GCT CCA CAA ATG    268
Ser Ser Gln Ser Pro Gln Gly Ser Ala Phe Arg Pro Ala Pro Gln Met
      10                      15                      20

ATT CAT TTC AGC CCC CTT GVS    289
Ile His Phe Ser Pro Leu Xaa
      25                      30

```

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 268 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 110..205
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.6

seq SLVPSMCFHVTNS/IK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

```

ACGCCAAATG ATTATACTCG GGACACCTGA CCCAGTTTCT TCAACAAATA TATGGAGGGA      60
AACGTGTGTG TAGGACGGGG GTRTTTTTAA TAAAAAGGGC TTTAAGACG ATG GAA ATG      118
                                     Met Glu Met
                                     -30

TTT GGT TWR RTT GAA AAA GAT TTT TCA TCA GTG GAA GGG GTT CTA TRG      166
Phe Gly Xaa Xaa Glu Lys Asp Phe Ser Ser Val Glu Gly Val Leu Xaa
      -25                      -20                      -15

AGC CTG GTA CCT TCA ATG TGT TTC CAT GTT ACC AAC TCC ATA AAG ATG      214
Ser Leu Val Pro Ser Met Cys Phe His Val Thr Asn Ser Ile Lys Met
      -10                      -5                      1

CCC TGG TTT CCC AGC CAA CCA GGT ACT TGC ACC CAG AAG GAT TGC CCT      262
Pro Trp Phe Pro Ser Gln Pro Gly Thr Cys Thr Gln Lys Asp Cys Pro
      5                      10                      15

CCG AAG
Pro Lys
20
268

```

(2) INFORMATION FOR SEQ ID NO: 119:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 148 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 5..142
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.6  
seq LLGVHASFQMSVA/AR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

```

AATT ATG CAG ATG CAC GGC TGG AGG TGG GAT CCA CAC AGC TCA GAA CAG      49
Met Gln Met His Gly Trp Arg Trp Asp Pro His Ser Ser Glu Gln
      -45                      -40                      -35

CTG GAT CTT GCT CAC ACT CTT TCA AGA GAA GCT TCC TTG GAG AAT AAC      97
Leu Asp Leu Ala His Thr Leu Ser Arg Glu Ala Ser Leu Glu Asn Asn
      -30                      -25                      -20

```

[illegible]

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 199 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymphocytes

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 89..190  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.6  
seq VGTGVLTSRLARA/TP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

TCGA	ACTTGG	AGATAGCTGG	TTCTCCTCGA	AATAGCTTTA	GGGCTAGCGT	GTAGTGTTAA	60
GTAGTGGTGG	TAGAGCACTG	AATATGGA	ATG GCC TCG CCT AGG GGT ACT GAC	112			
			Met Ala Ser Pro Arg Gly Thr Asp				
			-30				
TAT AAT CAA ACT CCG AAT ACC ACT ATG TAT TGC TAT GCA GTC GGA ACC	160						
Tyr Asn Gln Thr Pro Asn Thr Thr Met Tyr Cys Tyr Ala Val Gly Thr							
-25	-20	-15					
GGG GTG CTA ACG TCC CGG CTC GCG AGG GCA ACA CCC GGG	199						
Gly Val Leu Thr Ser Arg Leu Ala Arg Ala Thr Pro Gly							
-10	-5	1					

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 356 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens



(F) TISSUE TYPE: Lymphocytes

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 144..269
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.6  
seq LHCLCPFPALFLS/VT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

```

AAAAACTGCT ATGCTTCCGT TACTACTCCA GMTATACTGG CCTCTTGCTT TCSCYTCAGC    60
ATTACTAATA TATTCTCCC TTAGAGCCTT TGCCTGGTT TTTCCCTCTG CTAAAGCAT    120
TCCCCTCCCC ACCCCACATT TAC ATG GCT CCC ATC CTC AGC TCC TTC AAG TCT    173
                        Met Ala Pro Ile Leu Ser Ser Phe Lys Ser
                        -40                                -35

TTG CTC AAA TAT CAT CTT CTT GAG ACT TCT CTA AGC ATT CTA TTG AAA    221
Leu Leu Lys Tyr His Leu Leu Glu Thr Ser Leu Ser Ile Leu Leu Lys
                        -30                                -25                                -20

CCT GTA ACC TTG CAC TGC CTC TGC CCC TTT CCT GCT TTA TTT CTA TCT    269
Pro Val Thr Leu His Cys Leu Cys Pro Phe Pro Ala Leu Phe Leu Ser
                        -15                                -10                                -5

GTT ACA TTT ATC TAT CTG ACA TAT TAT ATA TTT AAC TTA TAT ATT TTG    317
Val Thr Phe Ile Tyr Leu Thr Tyr Tyr Ile Phe Asn Leu Tyr Ile Leu
      1                                5                                10                                15

TTT ATT GTC TGT CTC CTC TAC TGG AAT GTA CTT TCC ATG                    356
Phe Ile Val Cys Leu Leu Tyr Trp Asn Val Leu Ser Met
                        20                                25

```

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 155 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 63..128
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq ILVPWWLPFVYT/AI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

```

AGAAAAAAAAA CTTCAAATT GAGGATCTCT GTAATTTAAA TCCTACTTAA CAAAGTAAAT      60
GC ATG AAC AGG TTG TCT AAA CAT CTT ATT ATA CTT GTT CCT TGG TGG      107
  Met Asn Arg Leu Ser Lys His Leu Ile Ile Leu Val Pro Trp Trp
    -20                      -15                      -10

CTT CCT CCC TTT GTT TAC ACT GCC ATA TCC TAT GTC CAA CTC CCA GGG      155
Leu Pro Pro Phe Val Tyr Thr Ala Ile Ser Tyr Val Gln Leu Pro Gly
   -5                      1                      5

```

## (2) INFORMATION FOR SEQ ID NO: 123:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 217 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 122..208
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq ALITILILYSSNS/AI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

```

ACCTCTGCAC TCCTGCCCCG CCTGCCCCCG GCCTGTCTGC TGGAGGTGTG AACCCACATC      60
CCTGCCCCCA GGGCCACCTG CAGGACGCCG ACACCTACCC CTCAGCAGAC GCCGGAGAGA      120
A ATG AGT AGC AAC AAA GAG CAG CGG TCA GCA GTG TTC GTG ATC CTC TTT      169
  Met Ser Ser Asn Lys Glu Gln Arg Ser Ala Val Phe Val Ile Leu Phe
    -25                      -20                      -15

GCC CTC ATC ACC ATC CTC ATC CTC TAC AGC TCC AAC AGT GCC ATT GGG      217
Ala Leu Ile Thr Ile Leu Ile Leu Tyr Ser Ser Asn Ser Ala Ile Gly
   -10                      -5                      1

```

## (2) INFORMATION FOR SEQ ID NO: 124:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 367 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 149..352  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.5  
seq CLFLSPQSFVLVS/WA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

```

ATAACATTTG GTGCCGAAAG CCCGGGATAG GGGAACCTCT CCGGCAGACC TCTCCTCTAT    60
CCTCCCGGTA CCCACGTTCT CCCATGCAAG AGACTTCCCT CGCCCTCAGG ACCTCAGACC    120
AGCTCCGCGA GCACTCCGGC CTCTGTCT ATG GAT ATG AAA TCC AAC ACC GGT    172
                               Met Asp Met Lys Ser Asn Thr Gly
                               -65
CAC GGA CTC TTC TTG GGG AGA CAG CCT TCC TTC AGT GTT CGG TCA ATG    220
His Gly Leu Phe Leu Gly Arg Gln Pro Ser Phe Ser Val Arg Ser Met
-60                               -55                               -50                               -45
CCC GGG ACG CCC GCC TTG GCC ATT TGC CAG CCA CAC AAC CCA GGA CCT    268
Pro Gly Thr Pro Ala Leu Ala Ile Cys Gln Pro His Asn Pro Gly Pro
                               -40                               -35                               -30
CCA ATG GGG ACG CCC ACT GAG GAT CCT AGT GGT TGC TCT TTT CCT TGT    316
Pro Met Gly Thr Pro Thr Glu Asp Pro Ser Gly Cys Ser Phe Pro Cys
                               -25                               -20                               -15
CTC TTT CTT TCC CCC CAG TCT TTC CTT GTT CTA TCA TGG GCA ATT TCC    364
Leu Phe Leu Ser Pro Gln Ser Phe Leu Val Leu Ser Trp Ala Ile Ser
                               -10                               -5                               1
CGC                                                                    367
Arg
5

```

## (2) INFORMATION FOR SEQ ID NO: 125:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 300 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 193..279

(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.5  
seq LSLSSTLLLTSHH/HQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

```
ACAACGCACA CATCGTGAGA CGTTTTTCAA GACACCCGGC AGCCTTGGAG ACCCTGTCCT    60
GAAGAGAAGA GAAAGGAACC AGTCACGAAA CACCAGCTCG GCCCAGAGGA GACTAGAAAT    120
CCCCAGCGGC GCGCTGACT AACCTGCCGC TTTGCCAGGT GGGGGTGGGA TCAAACGCCC    180
TGAGAGTCCC GG ATG TCC GAG GCG GGA TGC AAA CCA TCC CGT CCT GAG CAC    231
      Met Ser Glu Ala Gly Cys Lys Pro Ser Arg Pro Glu His
                        -25                      -20

GGG TCC TTC CTC TCT CTT TCA TCC ACA CTT CTG TTA ACT TCC CAC CAC    279
Gly Ser Phe Leu Ser Leu Ser Ser Thr Leu Leu Thr Ser His His
      -15                      -10                      -5

CAT CAA TCA TCT GAT TTC GGG    300
His Gln Ser Ser Asp Phe Gly
  1                      5
```

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 422 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 255..284
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 201..230  
id R72787  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 121..152
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 90  
region 1..32  
id H73816  
est

(ix) FEATURE:

- (A) NAME/KEY: other

(B) LOCATION: 241..283  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 93  
                           region 161..203  
                           id SSC1D10  
                           est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 81..137  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 12.8  
                           seq XVFLVALLRGVQC/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

```

AAGCHCTGGG AGAGGAGCCC AGCACTAGAA GTCGGCGGWG TTTCCATTCG GTGATCAGCA      60
CTGAACACAG AGGACTCACC ATG GAG TCC GGG MWG GGG TGR GTT TTC CTC GTT      113
           Met Glu Ser Gly Xaa Gly Xaa Val Phe Leu Val
                        -15                               -10

GCT CTT TTA AGA GGT GTC CAG TGT CAG GTG CAG ATT GTG CAG TCT GGG      161
Ala Leu Leu Arg Gly Val Gln Cys Gln Val Gln Ile Val Gln Ser Gly
           -5                               1                               5

GGA GGC GTG GTC CAG CCT GGG AAG TCC CAG ACA CTC TCC TGT GTT ACC      209
Gly Gly Val Val Gln Pro Gly Lys Ser Gln Thr Leu Ser Cys Val Thr
           10                               15                               20

TAT GGA TTC AGA TTC GAT GAC TTT GGC TTC CAC TGG GTC CGC CAG GCT      257
Tyr Gly Phe Arg Phe Asp Asp Phe Gly Phe His Trp Val Arg Gln Ala
           25                               30                               35                               40

CCA GGC AAG GGG CTG GAA TGG GTG GCA ATG ATA CGT TAT GAT GGA AGT      305
Pro Gly Lys Gly Leu Glu Trp Val Ala Met Ile Arg Tyr Asp Gly Ser
           45                               50                               55

AAT AAA TTC TAC TCA AAG TCT GTT CAG GGC CGA TTT CTC ATC TCC AGA      353
Asn Lys Phe Tyr Ser Lys Ser Val Gln Gly Arg Phe Leu Ile Ser Arg
           60                               65                               70

GAC AAT TCC AGA AAC CAA GTC TAT TTG AGT CTG AAC AGA CTG AGA GTC      401
Asp Asn Ser Arg Asn Gln Val Tyr Leu Ser Leu Asn Arg Leu Arg Val
           75                               80                               85

GAC GAC ACG GCT GTC TAT TAT                                          422
Asp Asp Thr Ala Val Tyr Tyr
           90                               95
  
```

## (2) INFORMATION FOR SEQ ID NO: 127:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 366 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 134..363

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98  
region 1..230  
id AA009645  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: 289..339

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 9.3  
seq LFTLLLLQSLLLG/CC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

```
CAAATGATGT ATTCAAGTAA TAAAAGAATC CCTTTTATAA AATCTATTTT TCTTTAAATC   60
TTGGAAAAAT GTTGTTTT TAG CTCAGAGTGA TTTCAAAGTG GAATGCAACA GTAGTCAAGA  120
CTTGTGTACT ATAAATCCTT TTCTGATTCC TTACAGATTT GTAGTGATGA GGTTTAGATT  180
TAATTTTATA TATGGTKTAA ATAATTGTTA AGCKTATATA ACCTGATCTG AATTGCAGTT  240
GTTTGCACTT CCTCTATGAA AACTTCATTT ATCTAATAAG GAAGTCAA ATG CTT TGT   297
                                   Met Leu Cys
                                   -15

AGA CTA TTT ACC TTA CTT TTG TTG CAA TCA CTG TTG TTG GGT TGC TGT   345
Arg Leu Phe Thr Leu Leu Leu Leu Gln Ser Leu Leu Leu Gly Cys Cys
          -10                -5                1

ATA TAT ATR CCG GGC AAT GGG                                   366
Ile Tyr Xaa Pro Gly Asn Gly
          5
```

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 308 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 43..97  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 90  
                           region 1..55  
                           id H30111  
                           est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 18..95  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 9.3  
                           seq FLLLVAGPRWVLS/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

```

AGTGCTTTCT GAGAGTC ATG GAC CTC CTG CAC AAG AAC ATG AAA CAC CTG      50
      Met Asp Leu Leu His Lys Asn Met Lys His Leu
      -25                                -20

TGG TTC TTC CTC CTC CTG GTG GCA GGT CCC AGA TGG GTC CTG TCC CAA      98
Trp Phe Phe Leu Leu Leu Val Ala Gly Pro Arg Trp Val Leu Ser Gln
-15                                -5                                1

GTG CGG CTG GAA CAG TGG GGC TCT GGA CTA GTG AAG TCT TCC GAA ACG      146
Val Arg Leu Glu Gln Trp Gly Ser Gly Leu Val Lys Ser Ser Glu Thr
      5                                10                                15

CTG TCC CTC ACC TGC GCT GTC TAT GGT GGA TCC GCC ATC AGT GAC TAC      194
Leu Ser Leu Thr Cys Ala Val Tyr Gly Gly Ser Ala Ile Ser Asp Tyr
      20                                25                                30

TGG GCT TGG ATC CGT CAA TTC CCA GGA AAG GGA GTG GAG TGG ATC GGT      242
Trp Ala Trp Ile Arg Gln Phe Pro Gly Lys Gly Val Glu Trp Ile Gly
      35                                40                                45

GAA ATC AAT CAC AGT GGC GCC ACC CAC TAT ATC CGT CCC TCA GGG GTC      290
Glu Ile Asn His Ser Gly Ala Thr His Tyr Ile Arg Pro Ser Gly Val
      50                                55                                60                                65

GAG TCG CCA TCT CCG CTG      308
Glu Ser Pro Ser Pro Leu
      70

```

## (2) INFORMATION FOR SEQ ID NO: 129:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 355 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 41..237  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 1..197  
id R68856  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 236..354  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 197..315  
id R68856  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 44..354  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 1..311  
id T87538  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 43..331  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 2..290  
id W95563  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 62..237  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 1..176  
id H67429  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 236..354  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 176..294  
id H67429  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 65..354  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98



region 1..290  
id AA046628  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 125..277
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.1  
seq VCLCGTFCFPCLG/CQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

```

AAAGTTACCT CTCCCCTTTC ACGTARTTTT CATTTGTGGT GAGATTCTCT CCCARGCCAC      60
AAGACATTTT CTGCTCGGAA CCTTGTTTAC TAATTTCCAC TGCTTTTAAG GCCCTGCACT     120
GAAA ATG CAA GCT CAR GCG CCG GTG GTC GTT GTG ACC CAA CCT GGA GTC      169
  Met Gln Ala Gln Ala Pro Val Val Val Val Thr Gln Pro Gly Val
    -50                      -45                      -40

GGT CCC GGT CCG GCC CCC CAG AAC TCC AAC TGG CAG ACA GGC ATG TGT      217
Gly Pro Gly Pro Ala Pro Gln Asn Ser Asn Trp Gln Thr Gly Met Cys
  -35                      -30                      -25

GAC TGT TTC AGC GAC TGC GGA GTC TGT CTC TGT GGC ACA TTT TGT TTC      265
Asp Cys Phe Ser Asp Cys Gly Val Cys Leu Cys Gly Thr Phe Cys Phe
  -20                      -15                      -10                      -5

CCG TGC CTT GGG TGT CAA GTT GCA GCT GAT ATG AAT GAA TGC TGT CTG      313
Pro Cys Leu Gly Cys Gln Val Ala Ala Asp Met Asn Glu Cys Cys Leu
          1                      5                      10

TGT GGA ACA AGC GTC GCA ATG AGG ACT CTC TAM AGG AHC CGG              355
Cys Gly Thr Ser Val Ala Met Arg Thr Leu Xaa Arg Xaa Arg
    15                      20                      25

```

## (2) INFORMATION FOR SEQ ID NO: 130:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 229 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 177..229
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 90  
region 1..53  
id N41594

est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 74..127
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.7  
seq LLLLPVLGLLVSS/KT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

```

AAAGAAAGAG CTGCSGTGCA GGAATTCGTG TGCCGGATTT GGTTAGCTGA GCCCACCGAG      60
AGGCGCCTGC AGG ATG AAA GCT CTC TGT CTC CTC CTC CTC CCT GTC CTG      109
      Met Lys Ala Leu Cys Leu Leu Leu Leu Pro Val Leu
      -15                               -10

GGG CTG TTG GTG TCT AGC AAG ACC CTG TGC TCC ATG GAA GAA GCC ATC      157
Gly Leu Leu Val Ser Ser Lys Thr Leu Cys Ser Met Glu Glu Ala Ile
      -5                               1                               5                               10

AAT GAG AGG ATC CAG GAG GTC GCC GGC TCC CTA ATA TTT AGG GCA ATA      205
Asn Glu Arg Ile Gln Glu Val Ala Gly Ser Leu Ile Phe Arg Ala Ile
      15                               20                               25

AGC AGC ATT GGC CGA GGG AGC GAG
Ser Ser Ile Gly Arg Gly Ser Glu
      30

```

## (2) INFORMATION FOR SEQ ID NO: 131:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 265 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 87..259
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 82..254  
id AA075901  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 66..259
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 1..194

id R55519  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 66..259  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 2..195  
id H25630  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 67..259  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 3..195  
id H43485  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 78..259  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 1..182  
id H80718  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 89..151  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.6  
seq LLXIVGLXLPTXG/QX

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

```
ACCTGGCYSB MCMCTCCGCC TGGNCGCAGC AKCCACCGCH GCGTCCCTCT CTCCACGAGG    60
CTGCCGGCTT AGGACCCCCA KCTCCGAC ATG TCG CCC TCT GGT CGC CTG TGT    112
                               Met Ser Pro Ser Gly Arg Leu Cys
                               -20                               -15

CTT CTM AYC ATC GTT GGC CTR AWK CTC CCC ACC AKW GGA CAG RCG TTG    160
Leu Leu Xaa Ile Val Gly Leu Xaa Leu Pro Thr Xaa Gly Gln Xaa Leu
-10                               -5                               1

AAA GAT ACC RCG TCC AGT TCT TCA GCA GAC TCA ACT ATC ATG GAC ATT    208
Lys Asp Thr Xaa Ser Ser Ser Ser Ala Asp Ser Thr Ile Met Asp Ile
5                               10                               15

CAG GTC CCG ACA CGA GCC CCA GAT GCA GTC TAC ACA GAA CTC CAG CCC    256
Gln Val Pro Thr Arg Ala Pro Asp Ala Val Tyr Thr Glu Leu Gln Pro
20                               25                               30                               35

ACC CAC GGG    265
Thr His Gly
```

## (2) INFORMATION FOR SEQ ID NO: 132:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 314 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 79..240
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 28..189  
id AA122029  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 82..267
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 26..211  
id HUML1833  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 148..275
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..128  
id AA158721  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 147..209
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.2  
seq FLVSNMLLAAYG/SG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

```
AAGTTCTAAA GAGAGGCTGK WTACCAKAAC AGCATAACAA GGCAGGTCT GACTGCAAGG    60
CTGTGGACTG GGAGGCAGAG CCGCCGCCAA GGGGGCCTCG GTTAARCACT GGTCGTTCAA   120
TCACCTGCAA GACGAAGGAG GCAAGG ATG CTG TTG GCC TGG GTA CAA GCA TTC    173
               Met Leu Leu Ala Trp Val Gln Ala Phe
               -20                               -15
```

CTC	GTC	AGC	AAC	ATG	CTC	CTA	GCA	GAA	GCC	TAT	GGA	TCT	GGA	GGC	TGT	221
Leu	Val	Ser	Asn	Met	Leu	Leu	Ala	Glu	Ala	Tyr	Gly	Ser	Gly	Gly	Cys	
		-10					-5					1				
TTC	TGG	GAC	AAC	GGC	CAC	CTG	TAC	CGG	GAG	GAC	CAG	ACC	TCC	CCC	GCG	269
Phe	Trp	Asp	Asn	Gly	His	Leu	Tyr	Arg	Glu	Asp	Gln	Thr	Ser	Pro	Ala	
5					10				15						20	
CCG	GGC	CTC	CGC	TGC	CTC	AAC	TGG	CTG	GAC	GCG	CAG	AAC	GGG	CTG		314
Pro	Gly	Leu	Arg	Cys	Leu	Asn	Trp	Leu	Asp	Ala	Gln	Asn	Gly	Leu		
				25				30						35		

## (2) INFORMATION FOR SEQ ID NO: 133:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 421 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 67..123
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..57  
id H30111  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 41..137
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 90  
region 1..97  
id T27715  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 43..137
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 91  
region 2..96  
id T27727  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 136..213
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 93..170

id T27727  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 98..137  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 1..40  
id H43753  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 44..137  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 92  
region 4..97  
id T28164  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 26..211  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.8  
seq LXLTCVS VSGGSIS/RT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

AATTCANNTC CAACTCATAA GGGAA ATG CTT TCT GAG AGT CGT GGA CCT CCT	52
Met Leu Ser Glu Ser Arg Gly Pro Pro	
-60 -55	
GTG CAA GAA CAT GAA GCA CCT GTG GTT CTT CCT CCT GCT GGT GGC GGC	100
Val Gln Glu His Glu Ala Pro Val Val Leu Pro Pro Ala Gly Gly Gly	
-50 -45 -40	
TCC CAG ATG GGT CCT GTC CCA GCT GCA GMT GCA GGG GAG TCG GGC CCA	148
Ser Gln Met Gly Pro Val Pro Ala Ala Xaa Ala Gly Glu Ser Gly Pro	
-35 -30 -25	
GGA VTG GTG AAG CCT TTG GAG ACC CTG TSC CTC ACC TGC AGT GTC TCA	196
Gly Xaa Val Lys Pro Leu Glu Thr Leu Xaa Leu Thr Cys Ser Val Ser	
-20 -15 -10	
GGT GGC TCG ATC AGT AGG ACC AGT TTC TAC TGG GGC TGG ATC CGC CAG	244
Gly Gly Ser Ile Ser Arg Thr Ser Phe Tyr Trp Gly Trp Ile Arg Gln	
-5 1 5 10	
CCC CCA GGG AAG GGA CTG GAG TGG ATT GGG AGT ATC TAT GAT ARG GGC	292
Pro Pro Gly Lys Gly Leu Glu Trp Ile Gly Ser Ile Tyr Asp Xaa Gly	
15 20 25	
AGC ACC TAC TAC AAC CCG TCC CTC AGN NNN NRR GTC ACC ATT TCA GTA	340
Ser Thr Tyr Tyr Asn Pro Ser Leu Xaa Xaa Xaa Val Thr Ile Ser Val	
30 35 40	
GAC ACG TCC AAG AAC CAG GTG TCC CTG AAG GTG AGC TCT GTG ACC GCC	388
Asp Thr Ser Lys Asn Gln Val Ser Leu Lys Val Ser Ser Val Thr Ala	

45

50

55

GCG GAC ACG GCC GTA TAT CAC TGT GCG AGA GGG  
Ala Asp Thr Ala Val Tyr His Cys Ala Arg Gly  
60 65 70

421

## (2) INFORMATION FOR SEQ ID NO: 134:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 431 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 220..343
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..124  
id N57208  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 329..431
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93  
region 111..213  
id N57208  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 329..431
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 55..157  
id R94133  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 276..343
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..68  
id R94133  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 344..431

(C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 93  
 region 1..88  
 id AA110914  
 est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 21..344  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 5.3  
 seq ACMTLTASPGVFP/SL

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

AAACA	ACTCC	GGAAAGTACA	ATG	ACC	AGC	GGG	CAG	GCC	CGA	GCT	TCC	WYC	CAG	53
			Met	Thr	Ser	Gly	Gln	Ala	Arg	Ala	Ser	Xaa	Gln	
						-105							-100	
TCC	CCC	CAG	GCC	CTG	GAG	GAC	TCG	GGC	CCG	GTG	AAT	ATC	TCA	101
Ser	Pro	Gln	Ala	Leu	Glu	Asp	Ser	Gly	Pro	Val	Asn	Ile	Ser	
		-95					-90					-85		
ATC	ACC	CTA	ACC	CTG	GAC	CCA	CTG	AAA	CCC	TTC	GGA	GGG	TAT	149
Ile	Thr	Leu	Thr	Leu	Asp	Pro	Leu	Lys	Pro	Phe	Gly	Gly	Tyr	
	-80				-75						-70		Ser	
													Arg	
AAC	GTC	ACC	CAT	CTG	TAC	TCA	ACC	ATC	TTA	GGG	CAT	CAG	ATT	197
Asn	Val	Thr	His	Leu	Tyr	Ser	Thr	Ile	Leu	Gly	His	Gln	Ile	
-65					-60					-55			Gly	
													Leu	-50
TCA	GGC	AGG	GAA	GCC	CAC	GAG	GAG	ATA	AAC	ATC	ACC	TTC	ACC	245
Ser	Gly	Arg	Glu	Ala	His	Glu	Glu	Ile	Asn	Ile	Thr	Phe	Thr	
			-45					-40					Leu	
													Pro	-35
ACA	GCG	TGG	AGC	TCA	GAT	GAC	TGC	GCC	CTC	CAC	GGT	CAC	TGT	293
Thr	Ala	Trp	Ser	Ser	Asp	Asp	Cys	Ala	Leu	His	Gly	His	Cys	
		-30					-25						Glu	
													Gln	-20
GTG	GTA	TTC	ACA	GCC	TGC	ATG	ACC	CTC	ACG	GCC	AGC	CCT	GGG	341
Val	Val	Phe	Thr	Ala	Cys	Met	Thr	Leu	Thr	Ala	Ser	Pro	Gly	
	-15					-10					-5		Val	
													Phe	
CCG	TCA	CTG	TAC	AGC	CAC	CGC	ACT	GTG	TTC	CTG	ACA	CGT	ACA	389
Pro	Ser	Leu	Tyr	Ser	His	Arg	Thr	Val	Phe	Leu	Thr	Arg	Thr	
	1				5				10				Ala	
													Thr	15
CCA	CGC	TCT	GGT	ACA	AGA	TCT	TCA	CAA	CTG	CCA	GAG	ATG	CCA	431
Pro	Arg	Ser	Gly	Thr	Arg	Ser	Ser	Gln	Leu	Pro	Glu	Met	Pro	
			20					25						

## (2) INFORMATION FOR SEQ ID NO: 135:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 144 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR



(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 111..142  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 2..33  
id R30650  
est

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 1..63  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.1  
seq LLLKIWLLQRPES/QE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

ATG CTG GGA GGT GAC CAT AGG GCT CTG CTT TTA AAG ATA TGG CTG CTT	48
Met Leu Gly Gly Asp His Arg Ala Leu Leu Leu Lys Ile Trp Leu Leu	
-20 -15 -10	
CAA AGG CCA GAG TCA CAG GAA GGA CTT CTT CCA GGG AGA TTA GTG GTG	96
Gln Arg Pro Glu Ser Gln Glu Gly Leu Leu Pro Gly Arg Leu Val Val	
-5 1 5 10	
ATG GAG AGG AGA GTT AAA AAT GAC CTC ATG TCC TTC TTG TCC ACG GCG	144
Met Glu Arg Arg Val Lys Asn Asp Leu Met Ser Phe Leu Ser Thr Ala	
15 20 25	

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 301 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 248..300  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..53  
id HSC1XE021

est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 134..220
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.6  
seq SLMSLLDESSCQA/VG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

```

AAGCGTCCCT TTGTTGTGAA GGC GCCGGG CCTAGCGCTA TGCCTGCGGC GGAGACTGCA    60
TCAGGCTCTC GTCCTCGGGC TCCACCCAGG GAGCTGTGCC CAGACAGCAG AAGGGAAGGA    120
TGTCATTCT GAG ATG AGG TTC AGA AAG GCC TGG GCT CCT GTC CTG GCT    169
      Met Arg Phe Arg Lys Ala Trp Ala Pro Val Leu Ala
                        -25                      -20

GCT CTC TCC CAC TCC CTG ATG AGC TTG CTG GAT GAA AGC TCC TGT CAG    217
Ala Leu Ser His Ser Leu Met Ser Leu Leu Asp Glu Ser Ser Cys Gln
      -15                      -10                      -5

GCT GTG GGG CGT CCT GTG GAG AAA CTG GCA AGA AAC TGG TGG GGG CCC    265
Ala Val Gly Arg Pro Val Glu Lys Leu Ala Arg Asn Trp Trp Gly Pro
      1                      5                      10                      15

TTT CCA CCT ATA GCC AGC AAG GAA CTG AAC CCA GCG    301
Phe Pro Pro Ile Ala Ser Lys Glu Leu Asn Pro Ala
      20                      25

```

## (2) INFORMATION FOR SEQ ID NO: 137:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 377 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 85..377
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..293  
id AA148067  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 133..327
- (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98  
region 58..252  
id AA129289  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 75..136  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..62  
id AA129289  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 171..242  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 263..334  
id R56463  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 239..284  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 332..377  
id R56463  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 145..180  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 91  
region 236..271  
id R56463  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 63..308  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.4  
seq NLPHLQVVGLTWG/HI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

AACTTCCGGT CGCGCCASCG CCCGTTGCCA GTTCTGCGCG TGTCTGCAT CTCCAGTATG 60

GA ATG TAT GTD TGG CCC TGT GCT GTG GTC CTG GCC CAG TAC CTT TGG 107  
Met Tyr Val Trp Pro Cys Ala Val Val Leu Ala Gln Tyr Leu Trp  
-30 -75 -70

TTT CAC AGA AGA TCT CTG CCA GGC AAG GCC ATC TTA GAG ATT GGA GCT 155  
Phe His Arg Arg Ser Leu Pro Gly Lys Ala Ile Leu Glu Ile Gly Ala  
-65 -60 -55

GGA GTG AGC CTT CCA GGA ATT TTG GCT GCC AAA TGT GGT GCA GAA GTA	203
Gly Val Ser Leu Pro Gly Ile Leu Ala Ala Lys Cys Gly Ala Glu Val	
-50 -45 -40	
ATA CTG TCA GAC AGC TCA GAA CTG CCT CAC TGT CTG GAA GTC TGT CGG	251
Ile Leu Ser Asp Ser Ser Glu Leu Pro His Cys Leu Glu Val Cys Arg	
-35 -30 -25 -20	
CAA AGC TGC CAA ATG AAT AAC CTG CCA CAT CTG CAG GTG GTA GGA CTA	299
Gln Ser Cys Gln Met Asn Asn Leu Pro His Leu Gln Val Val Gly Leu	
-15 -10 -5	
ACA TGG GGT CAT ATA TCT TGG GAT CTT CTG GCT CTA CCA CCA CAA GAT	347
Thr Trp Gly His Ile Ser Trp Asp Leu Leu Ala Leu Pro Pro Gln Asp	
1 5 10	
ATT ATC CTT GCA TCT GAT GTG TTC TTT GAA	377
Ile Ile Leu Ala Ser Asp Val Phe Phe Glu	
15 20	

## (2) INFORMATION FOR SEQ ID NO: 138:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 380 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 308..380
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 29..101  
id W52747  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 265..294
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93  
region 1..30  
id W52747  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 3..170
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.4  
seq LWKLALQSSSCLS/LF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

AA ATG CTT AAT CCA GCA CAG TSC GAC ACT ATG CCT TGT GAA TAC CTC	47
Met Leu Asn Pro Ala Gln Xaa Asp Thr Met Pro Cys Glu Tyr Leu	
-55 -50 -45	
TCT TTG GAT GCA ATG GAA AAG TGG ATT ATC TTT GGC TTT ATT TTG TGC	95
Ser Leu Asp Ala Met Glu Lys Trp Ile Ile Phe Gly Phe Ile Leu Cys	
-40 -35 -30	
CAT GGG ATC CTA AAT ACT GRS GCT ACA GCA CTG AAC CTT TGG AAA CTA	143
His Gly Ile Leu Asn Thr Xaa Ala Thr Ala Leu Asn Leu Trp Lys Leu	
-25 -20 -15 -10	
GCT CTT CAA AGT AGC TCT TGC CTC TCT CTC TTT CGG GAT GAA GTT TTC	191
Ala Leu Gln Ser Ser Ser Cys Leu Ser Leu Phe Arg Asp Glu Val Phe	
-5 1 5	
CAC ATT CAC AAA GCT GCA GAA GAC TTA TTT GTA AAC ATA CGA GGC TAT	239
His Ile His Lys Ala Ala Glu Asp Leu Phe Val Asn Ile Arg Gly Tyr	
10 15 20	
AAT AAA CGT ATT AAT GAC ATA AGA GAA TGC AAG GRG GCA GCC GTG TCA	287
Asn Lys Arg Ile Asn Asp Ile Arg Glu Cys Lys Xaa Ala Ala Val Ser	
25 30 35	
CAT GCT GGT TCA ATG CAC AGA GAA AGA CGC AAG TVT TTA AGA TCT GCA	335
His Ala Gly Ser Met His Arg Glu Arg Arg Lys Xaa Leu Arg Ser Ala	
40 45 50 55	
CTG AAG GAA TTG GCT ACT GTC CTC TCT GAT CAA CCT GGA TTG CTA	380
Leu Lys Glu Leu Ala Thr Val Leu Ser Asp Gln Pro Gly Leu Leu	
60 65 70	

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 214 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 104..213
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 2..111  
id T11539  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 134..205  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.9  
seq GVCLSVPSLPSIS/RP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

```
ACACAGTGAT GTAGAGCTTT GGCTCTTTGT AACCAGGAGT TCGATAGGAA GACAACTTTG    60
AAAAAGCACT TTGTGACTGG CAGGGTGCCA TGCAGCCTCA GCTGTTTCATT TCCAAGGGTC    120
ATCCATTTAC AGG ATG AAT GCT CAA GCC TCT TCC TCC CGG TGC CAT GGA        169
           Met Asn Ala Gln Ala Ser Ser Ser Arg Cys His Gly
                   -20                               -15

GTC TGC CTG TCA GTC CCC TCC TTG CCC AGC ATC TCC CGC CCG CCG          214
Val Cys Leu Ser Val Pro Ser Leu Pro Ser Ile Ser Arg Pro Pro
      -10                -5                        1
```

(2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 485 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 263..411  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 155..303  
id AA005338  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 110..208  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 1..99  
id AA005338  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 409..469  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 302..362

id AA005338  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 206..250
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 98..142  
id AA005338  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 299..446
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 192..339  
id AA005431  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 110..208
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..99  
id AA005431  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 206..250
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 98..142  
id AA005431  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 263..300
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 155..192  
id AA005431  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 263..433
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 175..345  
id W78855  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 89..250
- (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99  
region 1..162  
id W78855  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 89..250  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 2..163  
id H52413  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 263..400  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 91  
region 176..313  
id H52413  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 263..389  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 5..131  
id AA057872  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 375..470  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 118..213  
id AA057872  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 264..464  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.8  
seq QVLDSVLVGPVPA/ER

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

```
ATGAACTCGG GTGCAGCCAA TCGAGGGCAA CGCTGCTACT TATCAGAGCA GAATGGGCTG 60
TAGTTTAGTG AAATAGGAAA GCTGCAACAC ACTGTGGAGT GCTCCCGTGT AAATAATAAG 120
AGGAAAAAAG TTTCTCAAST CGCCGCTGCA CGACGTCTGG CCGGCGCTGG AGCGGGGGTC 180
TGGGCTCTCC CGAGCGGCCG CGCGCTGGAC TTTATTGTGC CGCAACCAGC CCCAGTTCCC 240
ATTGTTTGTG TTTTTTTCAA AAT ATG GCA AAG GTT CAG GTG AAC AAT GTA GTG 293
Met Ala Lys Val Gln Val Asn Asn Val Val
```



-65

-60

GTG	CTG	GAT	AAC	CCT	TCT	CCT	TTC	TAC	AAC	CCG	TTC	CAG	TTC	GAG	ATC	341
Val	Leu	Asp	Asn	Pro	Ser	Pro	Phe	Tyr	Asn	Pro	Phe	Gln	Phe	Glu	Ile	
		-55					-50					-45				
ACC	TTC	GAG	TGC	ATC	GAG	GAC	CTG	TCT	GAA	GAC	TTG	GAA	TGG	AAA	ATT	389
Thr	Phe	Glu	Cys	Ile	Glu	Asp	Leu	Ser	Glu	Asp	Leu	Glu	Trp	Lys	Ile	
	-40					-35				-30						
ATC	TAT	GTG	GGC	TCT	GCA	GAA	AGT	GAA	GAA	TAC	GAT	CAA	GTT	TTA	GAC	437
Ile	Tyr	Val	Gly	Ser	Ala	Glu	Ser	Glu	Glu	Tyr	Asp	Gln	Val	Leu	Asp	
-25				-20					-15					-10		
TCT	GTT	TTA	GTG	GGT	CCT	GTT	CCC	GCA	GAA	AGG	CAT	ATG	TTT	GTA	TTT	485
Ser	Val	Leu	Val	Gly	Pro	Val	Pro	Ala	Glu	Arg	His	Met	Phe	Val	Phe	
			-5						1				5			

## (2) INFORMATION FOR SEQ ID NO: 141:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 169 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 58..160
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 48..150  
id AA037680  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 58..122
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 31..95  
id W81645  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 121..170
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 93..142  
id W81645  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 65..122  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 94  
 region 7..64  
 id W06951  
 est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 121..170  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 98  
 region 62..111  
 id W06951  
 est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 59..168  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 90  
 region 1..110  
 id AA034702  
 est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 56..118  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 3.7  
 seq ETCALASHSGSSG/SK

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

AAGCCTTCCG CCKTCCCCAA GCCAACGTCT CCGCCGTCGG CTCCGCGGCG CCGCC ATG	58
	Met
GCC GAC GTG GAA GAC GGA GAG GAA ACC TGC GCC CTG GCC TCT CAC TCC	106
Ala Asp Val Glu Asp Gly Glu Glu Thr Cys Ala Leu Ala Ser His Ser	
-20 -15 -10 -5	
GGG AGC TCA GGC TCC AAG TCG GGA GGC GAC AAG ATG TTC TCC CTC AAG	154
Gly Ser Ser Gly Ser Lys Ser Gly Gly Asp Lys Met Phe Ser Leu Lys	
1 5 10	
AAG TGG AAC GCG GTG	169
Lys Trp Asn Ala Val	
15	

## (2) INFORMATION FOR SEQ ID NO: 142:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 141 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia
- (ix) FEATURE:  
(A) NAME/KEY: other  
(B) LOCATION: 70..114  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 231..275  
id AA130776  
est
- (ix) FEATURE:  
(A) NAME/KEY: other  
(B) LOCATION: 88..139  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 1..52  
id AA121716  
est
- (ix) FEATURE:  
(A) NAME/KEY: other  
(B) LOCATION: 70..112  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 167..209  
id AA146672  
est
- (ix) FEATURE:  
(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 76..117  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.5  
seq WTCLLGDCGPPEA/FT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

```
ACWMTTCCCT CTGTTGGAGC TCAGAACATG ACACTCCAAA ACATGGCACC TTGGTAATTA    60
CGAAAACAGC AGAAA ATG TGG ACA TGC TTA CTC GGG GAT TGT GGC CCA CCA    111
      Met Trp Thr Cys Leu Leu Gly Asp Cys Gly Pro Pro
                        -10                               -5

GAA GCA TTT ACT TCC TAC CAA CCC CCC AGG                                141
Glu Ala Phe Thr Ser Tyr Gln Pro Pro Arg
      1                               5
```

(2) INFORMATION FOR SEQ ID NO: 143:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 247 base pairs  
(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 2..56  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 10..64  
id R86288  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 38..94  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 11.7  
seq VFCLLAVAPGAHS/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

ATCTCTACAG AACCTCTGA GAGGAAAGTT CTTACC ATG GAC TGG ACC TGG ASG	55
Met Asp Trp Thr Trp Xaa	
-15	
GTC TTC TGC TTG CTS GCT GTA GCT CCA GGT GCT CAC TCC CAG GTG CAA	103
Val Phe Cys Leu Leu Ala Val Ala Pro Gly Ala His Ser Gln Val Gln	
-10 -5 1	
CTG GTG CAG TCT SGG GCT SAG GTG AGG ASG CCT GGG GCC TCA GTG AAG	151
Leu Val Gln Ser Xaa Ala Xaa Val Arg Xaa Pro Gly Ala Ser Val Lys	
5 10 15	
GTT TCG TGC AAA CCA TCT GGA TAC AGT TTC ACC AGC CAC TAT GTS CAT	199
Val Ser Cys Lys Pro Ser Gly Tyr Ser Phe Thr Ser His Tyr Val His	
20 25 30 35	
TGG GTG CGG CAS GAC CCC GGA CAA SGG CTT GAG TGG ATG GGA GAT GGG	247
Trp Val Arg Xaa Asp Pro Gly Gln Xaa Leu Glu Trp Met Gly Asp Gly	
40 45 50	

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 307 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 117..250
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 17..150  
id R28399  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 240..278
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 141..179  
id R28399  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(179..305)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 59..185  
id R07794  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(149..184)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 181..216  
id R07794  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 187..305
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 19..137  
id H84735  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 195..234
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 250..289  
id R85011  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 240..278
- (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 94  
region 297..335  
id R85011  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 195..250  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 246..301  
id H52765  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 83..181  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 10.9  
seq LVSLLLLLTRVQP/GT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

AAACAGCTTC CCCTAGCACA GCACAGACTA CAGACAGGGG AGCCTCTGGG GCTGCAGACA	60
CTGCAGACGG ACGGACAGAC AG ATG GAC AAC AGC TGG AGG CTT GGC CCG GCC	112
Met Asp Asn Ser Trp Arg Leu Gly Pro Ala	
-30 -25	
ATA GGG CTC TCT GCG GGA CAG TCC CAG CTG CTA GTG TCG CTG TTG CTG	160
Ile Gly Leu Ser Ala Gly Gln Ser Gln Leu Leu Val Ser Leu Leu Leu	
-20 -15 -10	
CTA CTG ACC CGT GTC CAG CCT GGG ACA GAC GTG GCT GCC CCA GAG CAC	208
Leu Leu Thr Arg Val Gln Pro Gly Thr Asp Val Ala Ala Pro Glu His	
-5 1 5	
ATC AGC TAT GTG CCC CAG CTC TCA AAC GAC ACC TTG GCG GGG AGG CTC	256
Ile Ser Tyr Val Pro Gln Leu Ser Asn Asp Thr Leu Ala Gly Arg Leu	
10 15 20 25	
ACC CTG TCC ACC TTC ACG CTG GAG CAG CCT CTA GGC CAG TTC AGC AGC	304
Thr Leu Ser Thr Phe Thr Leu Glu Gln Pro Leu Gly Gln Phe Ser Ser	
30 35 40	
CGG	307
Arg	

## (2) INFORMATION FOR SEQ ID NO: 145:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 291 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 13..66  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 3..56  
id H30111  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 7..63  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 10.9  
seq FLLLVAAPRWVLS/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

AAGAAC ATG ARA CAC CTG TSG TTC TTC CTC CTG CTG GTG GCA GCT CCC	48
Met Xaa His Leu Xaa Phe Phe Leu Leu Leu Val Ala Ala Pro	
-15 -10	
AGA TGG GTC CTG TCC CAG GTG CTA CTA CAG GAG TCG GGC CCT GAA CTG	96
Arg Trp Val Leu Ser Gln Val Leu Leu Gln Glu Ser Gly Pro Glu Leu	
-5 1 5 10	
GTG AAG CCT TCA SAG ACC CTG TCC CTC ACC TGM GCT GTC TCT GGT GGC	144
Val Lys Pro Ser Xaa Thr Leu Ser Leu Thr Xaa Ala Val Ser Gly Gly	
15 20 25	
TCC ATC AGC GGT GGT CCT TAC TAT TGG AAC TGG GTC MGC CAG CAC CCA	192
Ser Ile Ser Gly Gly Pro Tyr Tyr Trp Asn Trp Val Xaa Gln His Pro	
30 35 40	
GGG AAG GGC CTG GAR WGG ATT GGC AAC ATC TAT TAC AAT GGG AGC ACC	240
Gly Lys Gly Leu Glu Xaa Ile Gly Asn Ile Tyr Tyr Asn Gly Ser Thr	
45 50 55	
TTC DHA RAA CCC GTC CCT CAA GAS TCG NCT TAT CAT ATC GYY AGA CGC	288
Phe Xaa Xaa Pro Val Pro Gln Xaa Ser Xaa Tyr His Ile Xaa Arg Arg	
60 65 70 75	
AGG	291
Arg	

## (2) INFORMATION FOR SEQ ID NO: 146:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 326 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 205..321  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 107..223  
id W05822  
est

## (ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 100..210  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 1..111  
id W05822  
est

## (ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: complement(218..248)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 105..135  
id AA135917  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 72..155  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 9.6  
seq LLTLLGLTEVAG/EE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

```

AGTCTTCBSG CAGGGCCTGA CATCTCCCA GAACAGACGT TTGAACAGAG CAGGCTTCTG    60
AGGTCTCCAA A ATG CCT GTC CCA GCC TCC TGG CCC CAT CCT CCT GGT CCT    110
      Met Pro Val Pro Ala Ser Trp Pro His Pro Pro Gly Pro
                -25                      -20

TTC CTG CTT CTG ACT CTA CTG CTR GGA CTT ACA GAA GTG GCA GGT GAG    158
Phe Leu Leu Leu Thr Leu Leu Leu Gly Leu Thr Glu Val Ala Gly Glu
-15                -10                -5                1

GAR GAG CTA CAG ATG ATT CAG CCT GAG AAG CTC CTG TTG GTC ACA GTT    206
Glu Glu Leu Gln Met Ile Gln Pro Glu Lys Leu Leu Leu Val Thr Val
                5                10                15

GGA AAG ACA GCC ACT CTG CAC TGC ACT GTG ACC TCC CTG CTT CCC GTG    254
Gly Lys Thr Ala Thr Leu His Cys Thr Val Thr Ser Leu Leu Pro Val
                20                25                30

GGA CCC CTC CTG TGG TTC AGA GGA GTT GGA CCA GGC CGG GAA TTA ATC    302

```



Gly Pro Val Leu Trp Phe Arg Gly Val Gly Pro Gly Arg Glu Leu Ile  
35 40 45

TAC AAT CAA AAA GAA GGC CTA MNB 326  
Tyr Asn Gln Lys Glu Gly Leu Xaa  
50 55

## (2) INFORMATION FOR SEQ ID NO: 147:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 330 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 70..216
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 68..214  
id W01897  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 4..56
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..53  
id W01897  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 237..273
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 238..274  
id W01897  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 177..331
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..155  
id H87389  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide

- (B) LOCATION: 217..261  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 9.6  
seq EYVLLLFLALCSA/KP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

```
AAGTTTATTC CAGTATCACC CAGGGTGCAG CCACACCAGG ACTGTGTTGA AGGGTGTTTT    60
TTTTCTTTTA AATGTAATAC CTCCTCATCT TTTCTTCTTA CACAGTGTCT GAGAACATTT    120
ACATTATAGA TAAGTAGTAC ATGGTGGATA ACTTCTACTT TTAGGAGGAC TACTCTCTTC    180
TGACAGTCCT AGACTGGTCT TCTACACTAA GACACC ATG AAG GAG TAT GTG CTC    234
                               Met Lys Glu Tyr Val Leu
                               -15                               -10

CTA TTA TTC CTG GCT TTG TGC TCT GCC AAA CCC TTC TTT AGC CCT TCA    282
Leu Leu Phe Leu Ala Leu Cys Ser Ala Lys Pro Phe Phe Ser Pro Ser
                               -5                               1                               5

CAC ATC GCA CTG AAG AAT ATG ATG CTG AAG GAT ATG GAA GAC ACA GAG    330
His Ile Ala Leu Lys Asn Met Met Leu Lys Asp Met Glu Asp Thr Glu
                               10                               15                               20
```

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 143 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 22..143  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 1..122  
id AA027314  
est

(ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 51..143  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 29..121  
id AA149456  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 69..143  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 92  
region 36..110  
id W17274  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 33..75  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..43  
id W17274  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 27..77  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 9.3  
seq LALSLILVLAFG/IP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

ACTCTTACTC ACCCTCTACC ACAGAC ATG GCT CAG TCA CTG GCT CTG AGC CTC	53
Met Ala Gln Ser Leu Ala Leu Ser Leu	
-15 -10	
CTT ATC CTG GTT CTG GCC TTT GGC ATC CCC AGG ACC CAA GGC AGT GAT	101
Leu Ile Leu Val Leu Ala Phe Gly Ile Pro Arg Thr Gln Gly Ser Asp	
-5 1 5	
GGA GGS GCT CAG GAC TGT TGC CTC AAG TAC AGC CAA ACG AGG	143
Gly Gly Ala Gln Asp Cys Cys Leu Lys Tyr Ser Gln Thr Arg	
10 15 20	

## (2) INFORMATION FOR SEQ ID NO: 149:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 305 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(E) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(228..303)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 206..281

id N70479  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(77..112)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 403..438  
id N70479  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(163..196)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 315..348  
id N70479  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 57..107  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 8.2  
seq LLLITAILAVAVG/FP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

```

AACTTGCCAT TTCTCATAAC AGCGTCAGAG AGAAAGAACT GACTGAAACG TTTGAG ATG      59
                                     Met
AAG AAA GTT CTC CTC CTG ATC ACA GCC ATC TTG GCA GTG GCT GTT GGT      107
Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val Gly
-15                               -10                               -5

TTC CCA GTC TCT CAA GAC CAA GAA CGA GAA AAA AGA AGT ATC AGT GAC      155
Phe Pro Val Ser Gln Asp Xaa Glu Arg Glu Lys Arg Ser Ile Ser Asp
1                               5                               10                               15

AGC GAT GAA TTA GCT TCA GGG TTT TTT GTG TTC CCT TAC CCA TAT CCA      203
Ser Asp Glu Leu Ala Ser Gly Phe Phe Val Phe Pro Tyr Pro Tyr Pro
20                               25                               30

TTT CGC CCA CTT CCA CCA ATT CCA TTT CCA AGA TTT CCA TGG TTT AGA      251
Phe Arg Pro Leu Pro Pro Ile Pro Phe Pro Arg Phe Pro Trp Phe Arg
35                               40                               45

CGT AAT TTT CCT ATT CCA ATA CCT GAA TCT GCC CCT ACA ACT CCC CTT      299
Arg Asn Phe Pro Ile Pro Ile Pro Glu Ser Ala Pro Thr Thr Pro Leu
50                               55                               60

CCG ATG
Pro Met
65

```

## (2) INFORMATION FOR SEQ ID NO: 150:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 323 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Lymph ganglia
- (ix) FEATURE:  
 (A) NAME/KEY: other  
 (B) LOCATION: complement(261..320)  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 95  
                                   region 222..281  
                                   id N70479  
                                   est
- (ix) FEATURE:  
 (A) NAME/KEY: other  
 (B) LOCATION: complement(110..145)  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 94  
                                   region 403..438  
                                   id N70479  
                                   est
- (ix) FEATURE:  
 (A) NAME/KEY: other  
 (B) LOCATION: complement(196..229)  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 94  
                                   region 315..348  
                                   id N70479  
                                   est
- (ix) FEATURE:  
 (A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 90..140  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 8.2  
                                   seq LLLITAILAVAVG/FP
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

```

AATATRARAC AGCTACAATA TTCCAGGGCC ARTCACTTGC CATTTCTCAT AACAGCGTCA      60
GAGAGAAAGA ACTGACTGAR ACGTTTGAG ATG AAG AAA GTT CTC CTC CTG ATC      113
                               Met Lys Lys Val Leu Leu Leu Ile
                               -15                               -10
ACA GCC ATC TTG GCA GTG GCT GTW GGT TTC CCA GTC TCT CAA GAC CAG      161
Thr Ala Ile Leu Ala Val Ala Val Gly Phe Pro Val Ser Gln Asp Gln
                               -5                               1                               5
GAA CGA GAA AAA AGA AGT ATC AGT GAC AGC GAT GAA TTA GCT TCA GGR      209

```

Glu	Arg	Glu	Lys	Arg	Ser	Ile	Ser	Asp	Ser	Asp	Glu	Leu	Ala	Ser	Gly		
		10					15					20					
WTT	TTT	GTG	TTC	CCT	TAC	CCA	TAT	CCA	TTT	CGC	CCA	CTT	CCA	CCA	ATT	257	
Xaa	Phe	Val	Phe	Pro	Tyr	Pro	Tyr	Pro	Phe	Arg	Pro	Leu	Pro	Pro	Ile		
	25					30				35							
CCA	TTT	CCA	AGA	TTT	CCA	TGG	TTT	AGA	CGT	AAW	TTT	CCK	ATT	CCA	ATA	305	
Pro	Phe	Pro	Arg	Phe	Pro	Trp	Phe	Arg	Arg	Xaa	Phe	Pro	Ile	Pro	Ile		
	40				45				50						55		
CCT	GAA	TCT	GCC	CCT	GGG											323	
Pro	Glu	Ser	Ala	Pro	Gly												
				60													

## (2) INFORMATION FOR SEQ ID NO: 151:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 302 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Placenta

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 130..267
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 104..241  
id T58582  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 28..129
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..102  
id T58582  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 18..270
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..253  
id C18397  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 97..300

(C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 99  
 region 2..205  
 id R62763  
 est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 97..300  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 99  
 region 1..204  
 id R63635  
 est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 150..203  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 8.1  
 seq LFTAILAFSLAQS/FG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

```

AAGCCTTCCT AGATCCCCTC CACTCGGTTT CTCTCTTTGC AGGAGCACCG GCAGCACCAG      60
TGTGTGAGGG GAGCAGGCAG CGGTCCTAGC CAGTTCCTTG ATCCTGCCAG ACCACCCAGC     120
CCCCGGCACA GAGCTGCTCC ACAGGCACC ATG AGG ATC ATG CTG CTA TTC ACA      173
                               Met Arg Ile Met Leu Leu Phe Thr
                               -15

GCC ATC CTG GCC TTC AGC CTA GCT CAG AGC TTT GGG GCT GTC TGT AAG      221
Ala Ile Leu Ala Phe Ser Leu Ala Gln Ser Phe Gly Ala Val Cys Lys
-10                               -5                               1                               5

GAG CCA CAG GAG GAG GTG GTT CCT GGC GGG GGC CGC AGC AAG AGG GAT      269
Glu Pro Gln Glu Glu Val Val Pro Gly Gly Gly Arg Ser Lys Arg Asp
          10          15          20

CCA GAT CTC TAC CAG CTG CTC CAG AGA CCC TGG      302
Pro Asp Leu Tyr Gln Leu Leu Gln Arg Pro Trp
          25          30

```

## (2) INFORMATION FOR SEQ ID NO: 152:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 416 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 17..102  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 90  
                           region 1..86  
                           id H16140  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 15..47  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 93  
                           region 16..48  
                           id H26913  
                           est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 42..92  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 7.9  
                           seq VLLLGLLSHCTVS/VS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

```

AAGTCTGGGC CTAWGGAAGC AGCACTGGTG GTGCCTCAGC C ATG GCC TGG ACC GTT      56
                                   Met Ala Trp Thr Val
                                   -15

CTC CTC CTC GGC CTC CTC TCT CAC TGC ACA GTG TCT GTG AGC TCC TAC      104
Leu Leu Leu Gly Leu Leu Ser His Cys Thr Val Ser Val Ser Ser Tyr
   -10                      -5                      1

GAA CTG ACT CAG CCA CCC TCA GTG TCA GTG GCC CCA GGA GAG ACG GCC      152
Glu Leu Thr Gln Pro Pro Ser Val Ser Val Ala Pro Gly Glu Thr Ala
   5                      10                      15                      20

ACC ATT TCC TGT GGG GCA AAC AAT GTT GGA AGA AAA AAT GTG CAG TGG      200
Thr Ile Ser Cys Gly Ala Asn Asn Val Gly Arg Lys Asn Val Gln Trp
                25                      30                      35

TAT CAG CAG AAG GCA GGC CAG GCC CCT GTG TTG GTC ATT TAC CAT GAT      248
Tyr Gln Gln Lys Ala Gly Gln Ala Pro Val Leu Val Ile Tyr His Asp
                40                      45                      50

GTC GAG CGG CCC TCG GGG ATT CCT GAG CGA TTC TCT GGC TCC AAC TCT      296
Val Glu Arg Pro Ser Gly Ile Pro Glu Arg Phe Ser Gly Ser Asn Ser
                55                      60                      65

GGG AGT CCG GCC AAA CTG ACC ATC AGC AGG GTC GAA GCC GGG GAT GAG      344
Gly Ser Pro Ala Lys Leu Thr Ile Ser Arg Val Glu Ala Gly Asp Glu
                70                      75                      80

GCC GAC TAT WAC TGT NAG GTG TGG GAC AGT GAC AGT GAT CAT ACG GTG      392
Ala Asp Tyr Xaa Cys Xaa Val Trp Asp Ser Asp Ser Asp His Thr Val
   85                      90                      95                      100

ATA TTC GGC GGC GGG ACC AAG CTG      416

```



Ile Phe Gly Gly Gly Thr Lys Leu  
105

(2) INFORMATION FOR SEQ ID NO: 153:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 519 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 166..454
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 91  
region 87..375  
id N23576  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 79..150
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..72  
id N23576  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 445..520
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 367..442  
id N23576  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(56..115)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 362..421  
id W15210  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(2..58)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 418..474

id W15210  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(2..118)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 18..134  
id R07597  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 274..447  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 7.7  
seq PXXXXLQTLPTASTX/XP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

```

AAGTACCTTT TCAGTCTTGC CCCAGAGGTT CCCTCAATTT CAGCAGCACC GAGCGGTTTA    60
TAATTCATTC AGTTTTCCAG GCCAGGCAGC CCGCTATCCT TGGATGGCCT TTCCACGCAA    120
TAGCATCATG CACTTGAACC ACACARCATA BGAGAGGGNR GNNNNSTAAT TTSTTGGACT    180
TSSVTSTBCC GCCACAGCRS AACACAGGTC TGGGAGGGAT CCCTGTAGCA GGTATTCCAG    240
CGTCTTCAGG AAACAGTTTA GACTCTCTTC AAG ATG ACA ATC CTC CAC ACT GGS      294
                        Met Thr Ile Leu His Thr Gly
                        -55

KAA AAT CCC TTC AGG CCC TCA CAG AGA TGG ACG GCC CCA GCG CTG CTC      342
Xaa Asn Pro Phe Arg Pro Ser Gln Arg Trp Thr Ala Pro Ala Leu Leu
-50                      -45                      -40

CAT CAC AGA CCC AMC ACA GBG CCC CCT TCA GKA CAC AGA TCC CGC TGC      390
His His Arg Pro Xaa Thr Xaa Pro Pro Ser Xaa His Arg Ser Arg Cys
-35                      -30                      -25                      -20

ACA GAG BYA GTT GGA ATC CCT RCS CTC CTC CTT CAR ACS CTT CCA GCT      438
Thr Glu Xaa Val Gly Ile Pro Xaa Leu Leu Leu Gln Thr Leu Pro Ala
-15                      -10                      -5

TCC ACT MCC CAS CCC CAG GCT TTC AGA CGG MCT TCA GAC CCC CCA GCA      486
Ser Thr Xaa Xaa Pro Gln Ala Phe Arg Arg Xaa Ser Asp Pro Pro Ala
1                      5                      10

PAA CCC CCA CAG ATT TAC TAC AGA GTT CAA CAC      519
Lys Pro Pro Gln Ile Tyr Tyr Arg Val Gln His
15                      20

```

## (2) INFORMATION FOR SEQ ID NO: 154:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 384 base pairs  
(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 21..73  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 92  
region 4..56  
id T28164  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 40..99  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 7.2  
seq LLLLVAAPKXXLS/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

ATATACTTTC TGAGAGTBCT GGACCTCCTG TGCAAGAAC ATG AAA CAC CTG TGG	54
Met Lys His Leu Trp	
-20	
TTC TTC CTC CTG CTG CTG GTG GCA GCT CCC AAA TKY KTC CTG TCC CAG	102
Phe Phe Leu Leu Leu Leu Val Ala Ala Pro Lys Xaa Xaa Leu Ser Gln	
-15 -10 -5 1	
GTG CAG CTG CGG GAG TCG GGC CCA GGA CTG GTG GAG CCT TCA CAG ACC	150
Val Gln Leu Arg Glu Ser Gly Pro Gly Leu Val Glu Pro Ser Gln Thr	
5 10 15	
CTG TCC CTC ACC TGC AGT GTT TCT CGT GGC TCC GTC AAC AGC GGG GGT	198
Leu Ser Leu Thr Cys Ser Val Ser Arg Gly Ser Val Asn Ser Gly Gly	
20 25 30	
TAC TAC TGG AGT TGG ATC CGC CAG CAC CCA GGA AAG GGC CTG GAG TGG	246
Tyr Tyr Trp Ser Trp Ile Arg Gln His Pro Gly Lys Gly Leu Glu Trp	
35 40 45	
ATT GGG TAC GTC TAT TAC GGT GGR RTA ACG TAC TAC AAC CCG TCC CTC	294
Ile Gly Tyr Val Tyr Tyr Gly Gly Xaa Thr Tyr Tyr Asn Pro Ser Leu	
50 55 60 65	
AAG AGT CGA GTC ACC CTA TCA GCA GAC ACG TCT AAG AAC CAG TTC TTC	342
Lys Ser Arg Val Thr Leu Ser Ala Asp Thr Ser Lys Asn Gln Phe Phe	
70 75 80	
CTG CGG CTG ACC TCA ATG ACT GCC GCG GAC ACG GCC AGT GGG	384
Leu Arg Leu Thr Ser Met Thr Ala Ala Asp Thr Ala Ser Gly	
85 90 95	

## (2) INFORMATION FOR SEQ ID NO: 155:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 69..194
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..126  
id R65867  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 208..245
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 140..177  
id R65867  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(208..245)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 194..231  
id R22927  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 120..182
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.9  
seq LVCGSLGLSNVSG/IY

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

```
AATATTTATA TAAATATTAB YTATTAGTGA GTTGTTATTC ATCCTTTGGA TGCAAATTGT    60
AAATTAGGAA ACTATTTTAT TACTGCTTTT TTGTGGTTAA ACACTTTATT TTAATATAA    119
ATG TTG AGT TAT TTT CTA TCC TCT TTG GTG TGC GGT AGT TTA GGT CTC    167
Met Leu Ser Tyr Phe Leu Ser Ser Leu Val Cys Gly Ser Leu Gly Leu
-20                               -15                               -10
AGT AAT GTT TCT GGT ATA TAT GAT TCC AAA AAA AAG CGA AAA ACA GGT    215
```

Ser Asn Val Ser Gly Ile Tyr Asp Ser Lys Lys Lys Arg Lys Thr Gly  
-5 1 5 10

GCT TTT AGG ACA CAA CTT TTC TGG GGA GTC GGG  
Ala Phe Arg Thr Gln Leu Phe Trp Gly Val Gly  
15 20

248

## (2) INFORMATION FOR SEQ ID NO: 156:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 380 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 184..354
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 172..342  
id AA043042  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 68..185
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 55..172  
id AA043042  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 12..52
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 1..41  
id AA043042  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 68..277
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 57..266  
id AA042861  
est

## (ix) FEATURE:

- (A) NAME/KEY: other

(B) LOCATION: 276..348  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 266..338  
id AA042861  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 10..52  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 1..43  
id AA042861  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 50..348  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 41..339  
id R76970  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 19..52  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 11..44  
id R76970  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 50..331  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 40..321  
id AA042849  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 10..52  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 1..43  
id AA042849  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 92..369  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 24..301  
id H60916  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 210..335
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.4  
seq ELPALALLHAGHA/EP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

```

AAGAGATCTG GGAACCCGGG AGCCGAGGTA ACGAACAGCT CGGTGGCAGG GCCTGACTGC      60
TGCGGAGGCC TCGGCAATAT TGATTTTAGA CAGGCAGACT TCTGCGTTAT GACCCGGCTG     120
CTGGGCTACG TGGACCCCCT GGATCCCAGC TTTGTGGCTG CCGTCATCAC CATCACCTTC     180
AATCCGCTCT ACTGGAATGT GGTTCACAG ATG GGA ACA CAA GAC CCG CAA GCT      233
                               Met Gly Thr Gln Asp Pro Gln Ala
                               -40                               -35
GAG CAG GGC CTT CGG ATC CCC CTA CCT GGC CTG CTA CTC TCT AAG CAT      281
Glu Gln Gly Leu Arg Ile Pro Leu Pro Gly Leu Leu Leu Ser Lys His
                               -30                               -25                               -20
CAC CAT CCT GCT CCT GAA CTT CCT GCG CTC GCA CTG CTT CAC GCA GGC      329
His His Pro Ala Pro Glu Leu Pro Ala Leu Ala Leu Leu His Ala Gly
                               -15                               -10                               -5
CAT GCT GAG CCA GCC CAG GAT GGA GAG CCT GGA CAC CCC CGC GGC CCA      377
His Ala Glu Pro Ala Gln Asp Gly Glu Pro Gly His Pro Arg Gly Pro
                               1                               5                               10
GGG                                                                    380
Gly
15

```

## (2) INFORMATION FOR SEQ ID NO: 157:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 23..251
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 78..306  
id T87972  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 251..283
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 307..339  
id T87972  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(36..283)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 89..336  
id AA040027  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(68..283)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 81..296  
id R82948  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(66..283)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 104..321  
id H26425  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(27..60)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 329..362  
id H26425  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(127..245)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 127..245  
id T55847  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(68..131)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 242..305  
id T55847



est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(237..283)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 88..134  
id T55847  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 61..159
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.3  
seq SXXPLXSVQLXHA/QR

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

```

ATAAATAGAA AAAAAAATT TTGTTTCCTA GGTGAAGGT CTAATTGATA CGTTTGACTT      60
ATG ATG ACC ATT TAT GCA CTT TCA AAT GAA TTT GCT TTC AAA ATA AAT      108
Met Met Thr Ile Tyr Ala Leu Ser Asn Glu Phe Ala Phe Lys Ile Asn
      -30                      -25                      -20

GAA GAG CAG CTG TCC TTM TTW CCW CTT TWA AGT GTY CAG CTG WGG CAT      156
Glu Glu Gln Leu Ser Xaa Xaa Pro Leu Xaa Ser Val Gln Leu Xaa His
      -15                      -10                      -5

GCT CAG AGG TTC CTG CTG GAT TCC AGC TGG AGC GGT GTG ATA CCC TTC      204
Ala Gln Arg Phe Leu Leu Asp Ser Ser Trp Ser Gly Val Ile Pro Phe
      1                      5                      10                      15

TTT TTC AGC TGT TCG TGC CTT CCT TTC TTG TAT CCA CCA AAG TGG AGA      252
Phe Phe Ser Cys Ser Cys Leu Pro Phe Leu Tyr Pro Pro Lys Trp Arg
      20                      25                      30

CAA ATA CAT GAT CTC AAA GAT ACA CAG TAC CGT TCG      288
Gln Ile His Asp Leu Lys Asp Thr Gln Tyr Arg Ser
      35                      40

```

## (2) INFORMATION FOR SEQ ID NO: 158:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 294 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: other

(B) LOCATION: 2..210  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 19..227  
id W04921  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 211..290  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 227..306  
id W04921  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(23..203)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 260..440  
id N70602  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(203..248)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 93  
region 216..261  
id N70602  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(251..290)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 175..214  
id N70602  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 47..177  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 1..131  
id W37690  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 211..290  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 165..244  
id W37690  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 175..210  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 97  
                           region 130..165  
                           id W37690  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 46..184  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 98  
                           region 1..139  
                           id W70167  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 228..290  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 96  
                           region 183..245  
                           id W70167  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 183..226  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 95  
                           region 139..182  
                           id W70167  
                           est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 217..279  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 6.1  
                           seq LEMLTAFASHIRA/RD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

```

ACCTTGGCTC GGCTTGGTCT GCGGCCTGTC AAACAGGTTC GGGTTCAGTT CTGTCCCTTC   60
GAGAAAAACG TGGAATCGAC GAGGACCTTC CTGCAGACGG TGAGCAGTGA GAAGGTCCGC   120
TCCACTAATC TCAACTGCTC AGTGATTGCG GACGTGAGGY ATGACGGCTC CGAGCCCTGC   180
GTGGACGTGC TGTTCGGAGA CGGGCATCGC CTGATT ATG CGC GGC GCT CAT CTC   234
                               Met Arg Gly Ala His Leu
                               -20
AYC GCT CTG GAA ATG CTC ACC GCC TTC GCC TCC CAC ATC CGG GCC AGG   282
Xaa Ala Leu Glu Met Leu Thr Ala Phe Ala Ser His Ile Arg Ala Arg
-15                               -5                               1
GAY GCG GCA AGG                                                    294

```

Asp Ala Ala Arg  
5

(2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(2..285)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 26..309  
id H54590  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 107..285
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 78..256  
id AA143123  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(169..285)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 320..436  
id AA142922  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(119..168)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 90  
region 436..485  
id AA142922  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 29..91
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93  
region 2..64

id N88917  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 83..140  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 93  
region 55..112  
id N88917  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 218..285  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 10..77  
id AA013161  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 76..231  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.1  
seq IILLIHTMQVCTT/HP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

```

AGTACCGATC CTCAGAGGAA GAGAAGAGAG TGACAGTCAT CAAAGCCCCG CATTACCCAG      60
GGATCGGGGCC CGTGG ATG AAT CCG GAA TCC CCA CAG CAA TTA GAA CGA CAG      111
      Met Asn Pro Glu Ser Pro Gln Gln Leu Glu Arg Gln
      -50                                -45

TCG ACC GGC CCA AGG ACT GGT ACA AGA CGA TGT TTA AGC AAA TTC ACA      159
Ser Thr Gly Pro Arg Thr Gly Thr Arg Arg Cys Leu Ser Lys Phe Thr
-40                                -35                                -30                                -25

TGG TGC ACA AGC CGG ATG ATG ACA CAG ACA TGT ATA ATA CTC CTT ATA      207
Trp Cys Thr Ser Arg Met Met Thr Gln Thr Cys Ile Ile Leu Leu Ile
      -20                                -15                                -10

CAT ACA ATG CAG GTC TGT ACA ACC CAC CCT ACA GTG CTC AGT CAC ACC      255
His Thr Met Gln Val Cys Thr Thr His Pro Thr Val Leu Ser His Thr
      -5                                1                                5

CTG CTG CAA AGA CCC AAA CCT ACA GAC CCC AGG      288
Leu Leu Gln Arg Pro Lys Pro Thr Asp Pro Arg
      10                                15

```

## (2) INFORMATION FOR SEQ ID NO: 160:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 286 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(117..284)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99  
region 40..207  
id H54590  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 120..284

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100  
region 78..242  
id AA143123  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(182..284)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100  
region 334..436  
id AA142922  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(132..181)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 90  
region 436..485  
id AA142922  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 231..284

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100  
region 10..63  
id AA013161  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 231..284

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100  
region 10..63  
id AA018245  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 188..244
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.1  
seq IILLIHTMQVCTT/HP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

```

AACTTCTCAT GAATGTTATG CTGTGTGTGG CAGAGAAAGA AAGCTGTTGA TGGGAGAGAA    60
TTACACAATC TGTTTTTCCA TTTGAATGAA ACATCATGAA CATTGCGATT TTGTAAATG    120
ACAGTCGACC GGCCCAAGGA CTGGTACAAG ACGATGTTTA AGCAAATTCA CATGGTGCAC    180
AAGCCGG ATG ATG ACA CAG ACA TGT ATA ATA CTC CTT ATA CAT ACA ATG    229
      Met Met Thr Gln Thr Cys Ile Ile Leu Leu Ile His Thr Met
                -15                      -10
CAG GTC TGT ACA ACC CAC CCT ACA GTG CTC AGT CAC ACC CTG CTG CAA    277
Gln Val Cys Thr Thr His Pro Thr Val Leu Ser His Thr Leu Leu Gln
  -5              1              5              10
AGA CCC ATG    286
Arg Pro Met

```

## (2) INFORMATION FOR SEQ ID NO: 161:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 355 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 176..321
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 158..303  
id R59094  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 101..176
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 82..157  
id R59094  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 56..104
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 38..86  
id R59094  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 177..256
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 193..272  
id R35689  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 101..171
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 117..187  
id R35689  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 60..104
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 77..121  
id R35689  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 101..171
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 115..185  
id H11787  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 39..104
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 54..119  
id H11787  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 2..31
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 17..46



id H11787  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 110..184
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6  
seq LLGLLVAVATVHL/VI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

```

AGCTTCGAAG GACGCCGCCG GGAGCTGCGG ACATGCGTGG AGTGGCAGTG CTAACGGCTG      60
GTGTCTCGCA CTGTTGGCCT GTGAAGGTAC GTGAAGCTGA AAGCCTGGA ATG GCT GGA      118
                                     Met Ala Gly
                                     -25

AAG GGG TCA TCA GGC AGG CGG CCC CTG CTG CTG GGG CTG CTG GTG GCC      166
Lys Gly Ser Ser Gly Arg Arg Pro Leu Leu Leu Gly Leu Leu Val Ala
   -20                      -15                      -10

GTA GCC ACT GTC CAC CTG GTC ATC TGT CCC TAC ACC AAA GTG GAG GAG      214
Val Ala Thr Val His Leu Val Ile Cys Pro Tyr Thr Lys Val Glu Glu
   -5                      1                      5                      10

AGC TTC AAC CTG CAG GCC ACA CAT GAC CTG CTC TAC CAC TGG CAA GAC      262
Ser Phe Asn Leu Gln Ala Thr His Asp Leu Leu Tyr His Trp Gln Asp
                15                      20                      25

CTG GAG CAG TAC GAC CAT CTT GAG TTC CCC GGA GTC GTC CCC AGG ACG      310
Leu Glu Gln Tyr Asp His Leu Glu Phe Pro Gly Val Val Pro Arg Thr
                30                      35                      40

TDC CTC GGG CCA GTG GTG ATC GCA GTG TTC TCC AGC CCC GCA GTG      355
Xaa Leu Gly Pro Val Val Ile Ala Val Phe Ser Ser Pro Ala Val
   45                      50                      55

```

## (2) INFORMATION FOR SEQ ID NO: 162:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 401 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 37..336
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 18..317

id H73135  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 336..384  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 93  
region 318..366  
id H73135  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 25..85  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..61  
id AA251602  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 30..95  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.9  
seq LIYILWQLTGSAA/SG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

ATTTCAGTGG CTGACTTCCA GAGAGCAAT ATG GCT GGT TCC CCA ACA TGC CTC	53
Met Ala Gly Ser Pro Thr Cys Leu	
-20 -15	
ACC CTC ATC TAT ATC CTT TGG CAG CTC ACA GGG TCA GCA GCC TCT GGA	101
Thr Leu Ile Tyr Ile Leu Trp Gln Leu Thr Gly Ser Ala Ala Ser Gly	
-10 -5 1	
CCC GTG AAA GAG CTG GTC GGT TCC GTT GGT GGG GCC GTG ACT TTC CCC	149
Pro Val Lys Glu Leu Val Gly Ser Val Gly Gly Ala Val Thr Phe Pro	
5 10 15	
CTG AAG TCC AAA GTA AAG CAA GTT GAC TCT ATT GTC TGG ACC TTC AAC	197
Leu Lys Ser Lys Val Lys Gln Val Asp Ser Ile Val Trp Thr Phe Asn	
20 25 30	
ACA ACC CCT CTT GTC ACC ATA CAG CCA GAA GGG GGC AHT ATC ATA GTG	245
Thr Thr Pro Leu Val Thr Ile Gln Pro Glu Gly Gly Xaa Ile Ile Val	
35 40 45 50	
ACC CAA AAT CGT AAT AGG GAG AGA GTA GAC TTC CCA GAT GGA GGC TAC	293
Thr Gln Asn Arg Asn Arg Glu Arg Val Asp Phe Pro Asp Gly Gly Tyr	
55 60 65	
TCC CTG AAG CTC AGC AAA CTG AAG AAG AAT GAC TCA GRN ATC TAC TAT	341
Ser Leu Lys Leu Ser Lys Leu Lys Lys Asn Asp Ser Xaa Ile Tyr Tyr	
70 75 80	
GTG GGG ATA TAC AGC TCA TCA CTC CAG CAG CCC TYC ACC CAG GAG TAC	389
Val Gly Ile Tyr Ser Ser Ser Leu Gln Gln Pro Xaa Thr Gln Glu Tyr	

85

90

95

GTG CTG CAT GTC  
Val Leu His Val  
100

401

## (2) INFORMATION FOR SEQ ID NO: 163:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 241 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 177..233
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 91  
region 314..370  
id T47889  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 179..232
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.8  
seq CFIILGLIICIQC/ST

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

AAAAGCCSSA CTTATTTTGG AAAGTTGTAG CCAGAAAAAT TAGAATTTAA TTTAAGCAGT 60  
AGAAAATAAT AAAAAGTGAA AAATGTTAGG CAACACTAGA ATTTAACAAC AGGTGTGCTA 120  
TGGTTTTTTTA AATATAATTT TCTTTTTCCTA GTTTCCTTATT TTTATTAAAA GACAAATC 178  
ATG GTA GGA ATG GTT TGC TTT ATT ATA CTT GGC TTA ATT ATT TGC ATA 226  
Met Val Gly Met Val Cys Phe Ile Ile Leu Gly Leu Ile Ile Cys Ile  
-15 -10 -5  
CAG TGC AGC ACG GGG 241  
Gln Cys Ser Thr Gly  
1

## (2) INFORMATION FOR SEQ ID NO: 164:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 352 base pairs

- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(5..325)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 10..330  
id W27422  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 22..250
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93  
region 151..379  
id AA153616  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 75..158
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 228..311  
id R19252  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 24..75
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 178..229  
id R19252  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 226..349
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..124  
id W03861  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 24..123
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 139..238

id AA029575  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 5..43
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.7  
seq MXLLHSLSSGVRA/PS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

ACAA ATG CBT CTT CTG CAC AGT CTA TCC TCG GGA GTT CGT GCA CCA TCT	49
Met Xaa Leu Leu His Ser Leu Ser Ser Gly Val Arg Ala Pro Ser	
-10 -5 1	
CCT GCC CCA TCA TCA GTA CCG TTA GGG TCA GAA AAG CCC AGC AAT GTG	97
Pro Ala Pro Ser Ser Val Pro Leu Gly Ser Glu Lys Pro Ser Asn Val	
5 10 15	
TCT CAG GAC AGG AAA GTT CCA GTC CCT ATT GGG ACT GAA CGT TCT GCA	145
Ser Gln Asp Arg Lys Val Pro Val Pro Ile Gly Thr Glu Arg Ser Ala	
20 25 30	
CGT ATC AGG CAA ACT GGA ACG TCA GCT CCA TCT GTT ATT GGG AGC AAT	193
Arg Ile Arg Gln Thr Gly Thr Ser Ala Pro Ser Val Ile Gly Ser Asn	
35 40 45 50	
TTG TCT ACA TCA GTA GGA CAT AGT GGC ATC TGG TCC TTT GAA GGG ATT	241
Leu Ser Thr Ser Val Gly His Ser Gly Ile Trp Ser Phe Glu Gly Ile	
55 60 65	
GGT GGC AAT CAA GAC AAA GTA GAC TGG TGT AAC CCT GGG ATG GGA AAT	289
Gly Gly Asn Gln Asp Lys Val Asp Trp Cys Asn Pro Gly Met Gly Asn	
70 75 80	
VCT ATG ATC CAC AGA CCG ATG TCT GAC CCA GGA GTA TTT TCA CAA CAT	337
Xaa Met Ile His Arg Pro Met Ser Asp Pro Gly Val Phe Ser Gln His	
85 90 95	
CAA GCA ACG GAK GCG	352
Gln Ala Thr Xaa Ala	
100	

## (2) INFORMATION FOR SEQ ID NO: 165:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 356 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 16..347
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..332  
id W56567  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 15..347
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..333  
id AA151004  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 15..296
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..282  
id AA147584  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 293..338
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 280..325  
id AA147584  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 53..296
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..244  
id W07033  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 293..337
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 242..286  
id W07033  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 55..338
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 1..284  
id H94668

est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 285..341
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.6  
seq PTLCVSSSPALWA/AS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

```

AACGCCTWTA AGACAGCGGA ACTAAGAAAA GAAGAGGCCT GTGGACAGAA CAATCATGTC    60
TGACTCCCTG GTGGTGTGCG AGGTAGACCC AGAGCTAACA GAAAAGCTGA KGAAATTCCG   120
CTTCCGAAAA GAGACAGACA ATGCAGCCAT CATAATGAAG GTGGACAAAG ACCGGCAGAT   180
GGTGGTGCTG GAGGAAGAAT TTCAGAACAT TTCCCCAGAG GVGCTCAAAA TGGAGTTGCC   240
GGAGAGACAG CCCAGGTTTCG TGGTTTACAG CTACAAGTAC GTGC ATG ACG ATG GCC   296
                               Met Thr Met Ala
GAG TGT CCT ACC CTT TGT GTT TCA TCT TCT CCA GCC CTG TGG GCT GCA   344
Glu Cys Pro Thr Leu Cys Val Ser Ser Ser Pro Ala Leu Trp Ala Ala
-15              -10              -5              1
AGC GAA ACA GGG                               356
Ser Glu Thr Gly
                    5

```

## (2) INFORMATION FOR SEQ ID NO: 166:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 463 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 182..352
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 179..349  
id AA057016  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 22..183
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97

region 20..181  
id AA057016  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 22..183  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 23..184  
id AA133917  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 182..293  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 182..293  
id AA133917  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 291..332  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 290..331  
id AA133917  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 182..308  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 119..245  
id R13065  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 80..183  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 18..121  
id R13065  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 65..139  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.5  
seq AQLFACLLRLGTQ/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

ATTTTTTGCT GCGCCAACGC ATGANCCGAA GCTCCGCTCA CGCCCGGCCT GATCCTGCCT 60



GAAG ATG GTG CCA CTG GTG GCT GTG GTA TCA GGG CCC CGT GCC CAG CTC	109
Met Val Pro Leu Val Ala Val Val Ser Gly Pro Arg Ala Gln Leu	
-25 -20 -15	
TTT GCC TGC CTG CTC AGG CTG GGC ACT CAG CAG GTC GGC CCC CTT CAG	157
Phe Ala Cys Leu Leu Arg Leu Gly Thr Gln Gln Val Gly Pro Leu Gln	
-10 -5 1 5	
CTG CAC ACC GGG GCC AGC CAT GCG GCC AGG AAC CAT TAT GAG GTG CTG	205
Leu His Thr Gly Ala Ser His Ala Ala Arg Asn His Tyr Glu Val Leu	
10 15 20	
GTG CTG GGT GGG GGC AGT GGC GGA ATC ACC ATG GCT GCC CGC ATG AAG	253
Val Leu Gly Gly Gly Ser Gly Gly Ile Thr Met Ala Ala Arg Met Lys	
25 30 35	
AGG AAA GTG GGT GCA GAG AAT GTG GCC ATT GTT GAG CCC AGT GAG AGA	301
Arg Lys Val Gly Ala Glu Asn Val Ala Ile Val Glu Pro Ser Glu Arg	
40 45 50	
CAT TTC TAC CAG CCA ATC TGG ACA CTG GTG GGT GCT GGT GCC AAA CAA	349
His Phe Tyr Gln Pro Ile Trp Thr Leu Val Gly Ala Gly Ala Lys Gln	
55 60 65 70	
TTG TCC TCA TCT GGT CGT CCC ACG GCA AGT GTG ATT CCA TCT GGT GTA	397
Leu Ser Ser Ser Gly Arg Pro Thr Ala Ser Val Ile Pro Ser Gly Val	
75 80 85	
GAA TGG ATC AAA GCT AGA GTG ACT GAG TTG AAC CAG ACA AGA CTG CAT	445
Glu Trp Ile Lys Ala Arg Val Thr Glu Leu Asn Gln Thr Arg Leu His	
90 95 100	
CAC ACA GAT GAC GAC GGA	463
His Thr Asp Asp Asp Gly	
105	

## (2) INFORMATION FOR SEQ ID NO: 167:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 370 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 34..282
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 29..277  
id HUM413F04B  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 272..363
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 268..359  
id HUM413F04B  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 27..181
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..155  
id R13204  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 174..282
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 147..255  
id R13204  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 275..344
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 249..318  
id R13204  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 27..282
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..256  
id R55599  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 275..344
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 250..319  
id R55599  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 27..344
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..318  
id H23092

est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 40..282  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 1..243  
id C05240  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 272..363  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 234..325  
id C05240  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 68..325  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.5  
seq ALLTGPTLGSSQA/RW

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

```
AACGAACGCA CGGCCGCGCA SATCTGTCTT GCTGGAACCTT TTTCCTAGAG GTTGAGCGGT      60
TTGCACA  ATG  TCG  GAA  ATG  GCT  GAG  TTG  TCC  GAG  CTG  TAT  GAA  GAG  AGC      109
      Met Ser Glu Met Ala Glu Leu Ser Glu Leu Tyr Glu Glu Ser
      -85                               -80                               -75
AGT  GAC  CTG  CAG  ATG  GAT  GTG  ATG  CCT  GGC  GAG  GGT  GAC  CTT  CCG  CAG      157
Ser Asp Leu Gln Met Asp Val Met Pro Gly Glu Gly Asp Leu Pro Gln
      -70                               -65                               -60
ATG  GAG  GTA  GGC  AGC  GGG  AGC  CGG  GAG  CTA  TCC  CTG  CGT  CCC  TCC  CGC      205
Met Glu Val Gly Ser Gly Ser Arg Glu Leu Ser Leu Arg Pro Ser Arg
      -55                               -50                               -45
AGC  GGG  GCC  CAA  CAG  CTC  GAG  GAG  GAA  GGC  CCA  ATG  GAG  GAG  GAG  GAG      253
Ser Gly Ala Gln Gln Leu Glu Glu Glu Gly Pro Met Glu Glu Glu Glu
      -40                               -35                               -30                               -25
GCC  CAG  CCA  ATG  GCG  SCG  CAG  AGG  GGA  AAC  GGA  GCC  TTG  CTA  ACG  GGC      301
Ala Gln Pro Met Ala Xaa Gln Arg Gly Asn Gly Ala Leu Leu Thr Gly
      -20                               -15                               -10
CCA  ACG  CTG  GGG  AGC  AGC  CAG  GCC  AGG  TGG  CGG  GCG  MAG  ACT  TCG  AGA      349
Pro Thr Leu Gly Ser Ser Gln Ala Arg Trp Arg Ala Xaa Thr Ser Arg
      -5                               1                               5
GCK  AGG  ACG  AGG  GCT  CCT  GGG
Ala Arg Thr Arg Ala Pro Gly
      10                               15
```

## (2) INFORMATION FOR SEQ ID NO: 168:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 354 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 51..195
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 38..182  
id W38899  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 197..324
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 183..310  
id W38899  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(207..349)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 141..283  
id W93646  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(58..195)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 455..592  
id W93646  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 47..195
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 32..180  
id W19506  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 197..324  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 181..308  
id W19506  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 197..338  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 194..335  
id W52820  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 71..195  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 69..193  
id W52820  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 65..195  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 9..139  
id W93906  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 207..269  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 309..371  
id W93906  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 244..288  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.3  
seq IVSVLALIPXTT/TT/LT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

```
AGAGCTGTNN CNSAAGTAGG GGAGGGCGGT GCTCCGCMGM GGTGGCGGDH TGCTATCGCT    60
TCGCAGAACC TACTCAGGCA GCCAGCTGAG AAGAGTTGAG GGAAAGTGCT GCTGCTGGGT    120
CTGACARACGC GATGGATAAC GTGCAGCCGA AAATAWAACA TCGMCCCTTC TGCTTCAGTG    180
TGAAARGCCA CGTGAYRGAW DCTGCGGCTG GATATTATCA ACTCACTGGT AACACAGTA    240
```

```

TTC ATG CTC ATC GTA TCT GTG TTG GCA CTG ATA CCA GAD ACC ACA ACA      288
Met Leu Ile Val Ser Val Leu Ala Leu Ile Pro Xaa Thr Thr Thr
-15                      -10                      -5

TTG ACM GTT GGT GGA GGG GTG TTT GCH HTT GTG ACA GCA GTA TGC TGT      336
Leu Thr Val Gly Gly Gly Val Phe Ala Xaa Val Thr Ala Val Cys Cys
1                      5                      10                      15

CTT GCC GAC GGG GGG GGG                                          354
Leu Ala Asp Gly Gly Gly
                20

```

## (2) INFORMATION FOR SEQ ID NO: 169:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 245 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 55..242
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 25..212  
id AA022775  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 55..242
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 23..210  
id R77353  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 55..242
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 16..203  
id W17384  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 55..242
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98

region 19..206  
id R05902  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 55..242  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 16..203  
id W76289  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 96..182  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5  
seq ELSLLPSSLWVLA/TS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

AAGAGACGTC ACCGGCTGCG CCCTTCAGTA TCGCGGACGG AAGATGGCGT CCGCCACCCG	60
TCTCATCCAG CGGCTGCGGA ACTGGGCGTC CGGGC ATG ACC TGC AGG GGA AGC	113
Met Thr Cys Arg Gly Ser	
-25	
TGC AGC TAC GCT ACC AGG AGA TCT CCA AGC GAA CTC AGC CTC CTC CCA	161
Cys Ser Tyr Ala Thr Arg Arg Ser Pro Ser Glu Leu Ser Leu Leu Pro	
-20 -15 -10	
AGC TCC CTG TGG GTC CTA GCC ACA AGC TCT CCA ACA ATT ACT ATT GCA	209
Ser Ser Leu Trp Val Leu Ala Thr Ser Ser Pro Thr Ile Thr Ile Ala	
-5 1 5	
CTC GCG ATG GCC GCC GGG AAT CTG TGC CCC CTT AGG	245
Leu Ala Met Ala Ala Gly Asn Leu Cys Pro Leu Arg	
10 15 20	

## (2) INFORMATION FOR SEQ ID NO: 170:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 222 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 20..220  
(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 97  
region 17..217  
id W24468  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 102..220  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 75..193  
id H71267  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 28..62  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..35  
id H71267  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 68..97  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 93  
region 40..69  
id H71267  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 45..220  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 3..178  
id W38688  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 45..220  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 3..178  
id W80906  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 52..220  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 1..169  
id AA037518  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide



- (B) LOCATION: 49..93
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.9  
seq AVVFVFSLLDCCA/LI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

```

AATCCGHNNN NGGTTTGACG GAAGGAGCGG CGGCGACGGA GGAGGAGG ATG GAG GCG      57
                                     Met Glu Ala
                                     -15

GTG GTG TTC GTC TTC TCT CTC CTC GAT TGT TGC GCG CTC ATC TTC CTC      105
Val Val Phe Val Phe Ser Leu Leu Asp Cys Cys Ala Leu Ile Phe Leu
      -10                               -5                               1

TCG GTC TAC TTC ATA ATT ACA TTG TCT RAT TTA GAA TGT GAT TAC ATT      153
Ser Val Tyr Phe Ile Ile Thr Leu Ser Xaa Leu Glu Cys Asp Tyr Ile
      5                               10                               15                               20

AAT GCT AGA TCA TGT TGC TCA AAA TTA AAC AAG TGG GTA ATT CCA GAA      201
Asn Ala Arg Ser Cys Cys Ser Lys Leu Asn Lys Trp Val Ile Pro Glu
      25                               30                               35

TTG ATT GGC CAT ACC ATT GGG
Leu Ile Gly His Thr Ile Gly
      40

```

(2) INFORMATION FOR SEQ ID NO: 171:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 249 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 14..197
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..184  
id AA043611  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 193..251
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 179..237  
id AA043611  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 79..168
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.9  
seq VAHALSLPAQSYG/ND

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

```

AGCCAGTWGA GAAGGACTCT GATCCGGCTC AGCTTTCCAA TCAGCTGCGG AAGGAGCCAC      60
GCTTTCGGGG GTTGCAAG ATG GCG GCC ACC AGT GGA ACT GAT GAG CCG GTT      111
      Met Ala Ala Thr Ser Gly Thr Asp Glu Pro Val
      -30                      -25                      -20

TCC GGG GAG TTG GTG TCW GTG GCA CAT GCG CTT TCT CTC CCA GCA CAG      159
Ser Gly Glu Leu Val Ser Val Ala His Ala Leu Ser Leu Pro Ala Gln
      -15                      -10                      -5

TCG TAT GGC AAC GAT CCT GAC ATT GAG ATG GCT TGG GCC ATG AGA GCA      207
Ser Tyr Gly Asn Asp Pro Asp Ile Glu Met Ala Trp Ala Met Arg Ala
      1                      5                      10

ATG CAG CAT GCT GAA GTC TAT TAC AAG CTG ATT TCA TCA GTT      249
Met Gln His Ala Glu Val Tyr Tyr Lys Leu Ile Ser Ser Val
      15                      20                      25

```

## (2) INFORMATION FOR SEQ ID NO: 172:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 406 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 212..349
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 158..295  
id W48792  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 54..181
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 2..129  
id W48792

est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 2..136  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 15..149  
id HUM031H01B  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 137..222  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 149..234  
id HUM031H01B  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 207..248  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 92  
region 220..261  
id HUM031H01B  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 275..325  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.6  
seq ALFLLLNEMVSG/VY

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

```
AATCTCAGCT GGTGGCTTT GGTAGAGCT CCCGTCAGAC TTTCGTTCCG CCCTAGGATT    60
GGTAGCCCC GAAGTGTGGG CTCTCTCCAG TACCAGACTC ATTTAGTAC CAGCCTTTGG    120
GAAGTCGTGT GAATACCTCG GTCTCTTAGC CACAGGGATA GAATGGCGGC CTGACGGAGC    180
CGCGGCGCCG GCGAAGTCGC TGAGGCGCGA GCTGGAACCC CCAGACCAGC TCAAACGGGA    240
GCCAAACTC GAAGCTTGGA AGAATTAGCA GGAA ATG GCG GAT GAG GCG TTG TTT    295
                               Met Ala Asp Glu Ala Leu Phe
                               -15

TTG CTT CTC CAT AAC GAG ATG GTG TCT GGA GTG TAC AAG TCC GCG GAS    343
Leu Leu Leu His Asn Glu Met Val Ser Gly Val Tyr Lys Ser Ala Xaa
-10          -5          1          5

ASG GGG AGG TGG AAA ACG GAC GAT GTA TTA CTA AGC TGG AAA ACA TGG    391
Xaa Gly Arg Trp Lys Thr Asp Asp Val Leu Leu Ser Trp Lys Thr Trp
          10          15          20

GGT TTC GAG TGG GAC    406
Gly Phe Glu Trp Asp
```

## (2) INFORMATION FOR SEQ ID NO: 173:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 399 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 16..307
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..292  
id H87111  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 92..352
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 2..262  
id W02272  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 173..400
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 2..229  
id W30926  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 68..198
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 20..150  
id R57641  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 22..378
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.3  
seq SVCLSIISMLSSC/KE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

ACTTGGGGGG ATGGTTCCAT C ATG GCG TCA ATG CAG AAA CGA CTA CAG AAA	51
Met Ala Ser Met Gln Lys Arg Leu Gln Lys	
-115 -110	
GAA CTG TTG GCT TTG CAA AAT GAC CCA CCT CCT GGA ATG ACC TTA AAT	99
Glu Leu Leu Ala Leu Gln Asn Asp Pro Pro Pro Gly Met Thr Leu Asn	
-105 -100 -95	
GAG AAG AGT GTT CAA AAT TCA ATT ACA CAG TGG ATT GTA GAC ATG GAA	147
Glu Lys Ser Val Gln Asn Ser Ile Thr Gln Trp Ile Val Asp Met Glu	
-90 -85 -80	
GGT GCA CCA GGT ACC TTA TAT GAA GGG GAA AAA TTT CAA CTT CTA TTT	195
Gly Ala Pro Gly Thr Leu Tyr Glu Gly Glu Lys Phe Gln Leu Leu Phe	
-75 -70 -65	
AAA TTT AGT AGT CGA TAT CCT TTT GAC TCT CCT CAG GTC ATG TTT ACT	243
Lys Phe Ser Ser Arg Tyr Pro Phe Asp Ser Pro Gln Val Met Phe Thr	
-60 -55 -50	
GGT GAA AAT ATT CCT GTT CAT CCT CAT GTT TAT AGC AAT GGT CAT ATC	291
Gly Glu Asn Ile Pro Val His Pro His Val Tyr Ser Asn Gly His Ile	
-45 -40 -35 -30	
TGT TTA TCC ATT CTA ACA GAA GAC TGG TCC CCA GCG CTC TCA GTC CAA	339
Cys Leu Ser Ile Leu Thr Glu Asp Trp Ser Pro Ala Leu Ser Val Gln	
-25 -20 -15	
TCA GTT TGT CTT AGC ATT ATT AGC ATG CTT TCC AGC TGC AAG GAA AAG	387
Ser Val Cys Leu Ser Ile Ile Ser Met Leu Ser Ser Cys Lys Glu Lys	
-10 -5 1	
AGA CGA CCA CCG	399
Arg Arg Pro Pro	
5	

(2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 425 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 265..426
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 57..218

id AA083634  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 262..420  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 91  
region 63..221  
id W71503  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 265..420  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 91  
region 193..348  
id W55411  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 249..329  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.1  
seq VLMFCVTPPELET/KX

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

```

AAGAACATCA GAAGTATATC TACATGAAGA ATTACAGCAA GACATGCAAA AGTTTAAGAA    60
TGAGGTCAAC ACATTAGAAG AAGAGTTCCT GGCTTTGAAG AAAGAAAATG TTCAACTTCA    120
TAAAGAGGTT GAAGAAGAAA TGGAGAAGCA CAGAAGTAAT AGCACAGAAT TATCAGGAAC    180
CCTAACTGAT GGTACTACTG TTGGCAATGA TGATGATGGA CTAAATCAGC AGATTCCTAG    240
GAAGGAAA ATG AAG ADM ATG ACA GGC TCT GAA AAT TGG AAA ACC AAG AAG    290
      Met Lys Xaa Met Thr Gly Ser Glu Asn Trp Lys Thr Lys Lys
      -25                                -20                                -15

GTT TTG ATG TTT TGT GTG ACG CCA CCT GAA TTA GAA ACC AAG RTG AAC    338
Val Leu Met Phe Cys Val Thr Pro Pro Glu Leu Glu Thr Lys Xaa Asn
      -10                                -5                                1

ATA ACC AAA GGT GGT CTG GTG TTG TTT WCA GCA AAC TCG AAT TCA TCA    386
Ile Thr Lys Gly Gly Leu Val Leu Phe Xaa Ala Asn Ser Asn Ser Ser
      5                                10                                15

TGT ATG GAG CTA TCA AAG AAA ATT GCA GAG CGG CCA GCG    425
Cys Met Glu Leu Ser Lys Lys Ile Ala Glu Arg Pro Ala
      20                                25                                30

```

## (2) INFORMATION FOR SEQ ID NO: 175:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 176 base pairs

- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 18..170
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 45..197  
id AA102765  
est

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 60..158
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.1  
seq LFMRTLCSPPGPS/QP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

AAGACGATTG GTCGGGCCAC GCCAGATCTC AGGATGATGG GGCGCACCTG GGGTTTGCC	59
ATG CAG CAC ATC GTG GGT GTG CCC CAC GTA CTG GTT CGG AGG GGC CTC	107
Met Gln His Ile Val Gly Val Pro His Val Leu Val Arg Arg Gly Leu	
-30 -25 -20	
CTT GGA AGG GAC CTC TTC ATG ACC AGG ACT CTC TGC AGC CCA GGC CCA	155
Leu Gly Arg Asp Leu Phe Met Thr Arg Thr Leu Cys Ser Pro Gly Pro	
-15 -10 -5	
AGC CAG CCC AGA GAG GCC GGG	176
Ser Gln Pro Arg Glu Ala Gly	
I 5	

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 302 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: other

(B) LOCATION: 33..229  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 2..198  
id W87850  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 209..298  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 179..268  
id W87850  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 200..298  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..99  
id AA043526  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 227..298  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..72  
id H68037  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 235..298  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 3..66  
id N78231  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 99..155  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.8  
seq ALALASSQSHLLG/RD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

AATTAAGTTC CGTTGTGGTT CACTCTGGTA TATCCTCTGA AAGTGGGCAT TACTATTCTT 60

ATGCCAGAAA TATCACAAGT ACAGACTCTT CATATCAG ATG TAC CAC CAG TCT GAG 116  
Met Tyr His Gln Ser Glu  
-15

GCT CTG GCA TTA GCA TCC TCC CAG AGT CAT TTA CTA GGG AGA GAT AGT 164  
Ala Leu Ala Leu Ala Ser Ser Gln Ser His Leu Leu Gly Arg Asp Ser



-10

-5

1

CCC	AGT	GCA	GTT	TTT	GAA	CAG	GAT	TTG	GAA	AAT	AAG	GAA	ATG	TCA	AAA	212
Pro	Ser	Ala	Val	Phe	Glu	Gln	Asp	Leu	Glu	Asn	Lys	Glu	Met	Ser	Lys	
	5					10					15					
GAA	TGG	TTT	TTA	TTT	AAT	GAC	AGT	AGA	GTG	ACA	TTT	ACT	TCA	TTT	CAG	260
Glu	Trp	Phe	Leu	Phe	Asn	Asp	Ser	Arg	Val	Thr	Phe	Thr	Ser	Phe	Gln	
	20				25				30						35	
TCA	GTC	CAG	AAA	ATT	ACG	AGC	AGG	TTT	CCA	AAG	GAC	ACC	TGG			302
Ser	Val	Gln	Lys	Ile	Thr	Ser	Arg	Phe	Pro	Lys	Asp	Thr	Trp			
				40					45							

## (2) INFORMATION FOR SEQ ID NO: 177:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(61..133)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 46..118  
id R15459  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(142..180)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..39  
id R15459  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 52..117
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq FASVAMICAIASG/SE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

AATGGGGGCC	GTGTAGCTGG	CCTTCTGCCT	GCCAGCTACA	CGGCCCCCAT	C	ATG	AGT	57
						Met	Ser	

GGC CAG GGC CTA GCA GGC TTC TTT GCC TCC GTG GCC ATG ATC TGC GCT	105
Gly Gln Gly Leu Ala Gly Phe Phe Ala Ser Val Ala Met Ile Cys Ala	
-20 -15 -10 -5	
ATT GCC AGT GGC TCG GAG CTA TCA GAA AGT GCC BAN GGC TAC TTT ATC	153
Ile Ala Ser Gly Ser Glu Leu Ser Glu Ser Ala Xaa Gly Tyr Phe Ile	
1 5 10	
ACA GCC TGT GCT GTT ATC ATT TTG ACC ATC ATC TGT TAC CTG GGC CTG	201
Thr Ala Cys Ala Val Ile Ile Leu Thr Ile Ile Cys Tyr Leu Gly Leu	
15 20 25	
CCC CGC CAA GGG	213
Pro Arg Gln Gly	
30	

## (2) INFORMATION FOR SEQ ID NO: 178:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 264 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(15..226)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..212  
id N56211  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 125..198
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 87..160  
id W37072  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 72..127
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 35..90  
id W37072  
est

## (ix) FEATURE:

- (A) NAME/KEY: other

(B) LOCATION: 191..245  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 92  
region 152..206  
id W37072  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 38..73  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 2..37  
id W37072  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 125..249  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 87..211  
id W37073  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 72..124  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 35..87  
id W37073  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 38..73  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 2..37  
id W37073  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 52..235  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 1..184  
id T34830  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 52..127  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 1..76  
id AA069883  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 125..198  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 73..146  
id AA069888  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 191..243  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 138..190  
id AA069888  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 115..204  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.7  
seq SMMLLTVYGGYLC/SV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

```
ACGAGTGCTG CGTTCGGCTG TGCTGGGAAG TTGCGTAGAC AGTGGCCTCG AGACCCTGCC      60
TGCCTGAGGA GGCCTCGGTT GGATGCGAAG GAGCTGCAGC ATCCAGGGGA CAAG ATG      117
                                     Met
                                     -30
CCA ACT GGC AAG CAG CTA GCT GAC ATT GGC TAT AAG ACC TTC TCT ACC      165
Pro Thr Gly Lys Gln Leu Ala Asp Ile Gly Tyr Lys Thr Phe Ser Thr
               -25                -20                -15
TCC ATG ATG CTT CTC ACT GTG TAT GGG GGG TAC CTC TGC AGT GTC CGA      213
Ser Met Met Leu Leu Thr Val Tyr Gly Gly Tyr Leu Cys Ser Val Arg
               -10                -5                1
GTC TAC CAC TAT TTC CAG TGG CGC AGG GCC CAG CGC CAS GCC GMA GAA      261
Val Tyr His Tyr Phe Gln Trp Arg Arg Ala Gln Arg Xaa Ala Xaa Glu
               5                10                15
GGG
Gly
20
```

## (2) INFORMATION FOR SEQ ID NO: 179:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 97 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(2..95)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 231..324  
id N32226  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(2..85)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 243..326  
id N32240  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 44..85
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq FVPCLTVTAAVCG/XX

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

```
AACACATTTTC TGTATCGCCC CGTGAAGAGC TCTACGCACA TGA ATG TTC CCC GTC      55
                                   Met Phe Pro Val

TGT TTA ACT GTT ACG GCM GCA GTG TGT GGG CNG CAS GCA CAG      97
Cys Leu Thr Val Thr Ala Ala Val Cys Gly Xaa Xaa Ala Gln
-10                      -5                      1
```

## (2) INFORMATION FOR SEQ ID NO: 180:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 241 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 60..240
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97

region 66..246  
id R89543  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 60..223  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 66..229  
id H59647  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 49..205  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 34..190  
id N34164  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 95..139  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.5  
seq VIFFACVVRVRDG/LP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

```

AAAGATTGCT GAGGAGGCGG CGGGTAGCTG GCAGGCGCCG ACTTCCGAAN GCCGCCGTCC      60
GGGCGAGGTG TCCTCATGAC TTCTCTTG TG GACC ATG TCC GTG ATC TTT TTT GCC      115
                               Met Ser Val Ile Phe Phe Ala
                               -15                      -10

TGT GTG GTA CGG GTA AGG GAT GGA CTG CCC CTC TCA GCC TCT ACT GAT      163
Cys Val Val Arg Val Arg Asp Gly Leu Pro Leu Ser Ala Ser Thr Asp
          -5                      1                      5

TTT TAC CAC ACC CAA GAT TTT TTG GAA TGG AGG AGA CGG CTC AAG AGT      211
Phe Tyr His Thr Gln Asp Phe Leu Glu Trp Arg Arg Arg Leu Lys Ser
      10                      15                      20

TTA GCC TTG CGA CTG GCC CAG TAT CCA GGG      241
Leu Ala Leu Arg Leu Ala Gln Tyr Pro Gly
      25                      30

```

## (2) INFORMATION FOR SEQ ID NO: 181:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 316 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 80..256
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 68..244  
id H46779  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 249..312
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 238..301  
id H46779  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 80..312
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 68..300  
id H46081  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 73..293
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93  
region 2..222  
id W07846  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 262..312
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 192..242  
id W07846  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 80..283
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 12..215  
id AA022743  
est

## (ix) FEATURE:

- (A) NAME/KEY: other

(B) LOCATION: 278..312  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 256..290  
id AA022743  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 50..253  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.5  
seq LVLDVVMLLLYLGI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

ACCCTGTTTC	CGGCAGCGCS	CGCTGCTCCG	GGAGCCGCTG	TGGCAGCGT	ATG CTS AVG	58
					Met Leu Xaa	
GGG GGA CTG AAG ATG GCG CCG CGA GGT AAA CGG TTG TCC TCC ACC CCG	106					
Gly Gly Leu Lys Met Ala Pro Arg Gly Lys Arg Leu Ser Ser Thr Pro						
-65 -60 -55 -50						
CTG GAA ATC CTG TTC TTT CTG AAC GGG TGG TAT AAT GCT ACC TAT TTC	154					
Leu Glu Ile Leu Phe Phe Leu Asn Gly Trp Tyr Asn Ala Thr Tyr Phe						
-45 -40 -35						
CTG CTG GAA CTT TTC ATA TTT CTG TAT AAA GGT GTC CTG CTA CCA TAT	202					
Leu Leu Glu Leu Phe Ile Phe Leu Tyr Lys Gly Val Leu Leu Pro Tyr						
-30 -25 -20						
CCA ACA GCT AAC CTA GTA CTG GAT GTG GTG ATG CTC CTC CTT TAT CTT	250					
Pro Thr Ala Asn Leu Val Leu Asp Val Val Met Leu Leu Leu Tyr Leu						
-15 -10 -5						
GGA ATT GAA GTA ATT CGC CTG TTT TTT GGT ACA AAG GGA AAC CTC TGC	298					
Gly Ile Glu Val Ile Arg Leu Phe Phe Gly Thr Lys Gly Asn Leu Cys						
1 5 10 15						
CAG CGA AAG ATG CCG AGG	316					
Gln Arg Lys Met Pro Arg						
20						

(2) INFORMATION FOR SEQ ID NO: 182:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 292 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(11) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganqlia

(ix) FEATURE:



(A) NAME/KEY: other  
 (B) LOCATION: complement(81..293)  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 97  
                           region 33..245  
                           id HSC1JB122  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: complement(113..289)  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 97  
                           region 65..241  
                           id H05810  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: complement(131..289)  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 98  
                           region 37..195  
                           id HSC20F042  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 245..289  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 100  
                           region 1..45  
                           id H14940  
                           est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 98..205  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 3.5  
                           seq PALTILHLPGTEG/VA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

```

AACTCTCTGG CCTGTGTCTA GTTGTITGAT TCAGACAGCT GCCTGGGATC CCTCATCCTC   60
ATACCCACCC CCACCCAAGG GCCTGGCCTG AGCTGGG ATG ATT GGA GGG GGG AGG   115
                               Met Ile Gly Gly Gly Arg
                               -35

TGG GAT CCT CCA GGT GCA CAA GCT CCA AGC TCC CAG GCA TTC CCC AGG   163
Trp Asp Pro Pro Gly Ala Gln Ala Pro Ser Ser Gln Ala Phe Pro Arg
-30          -25          -20          -15

AGG CCA GCC TTG ACC ATT CTC CAC CTG CCA GGG ACA GAG GGG GTG GCC   211
Arg Pro Ala Leu Thr Ile Leu His Leu Pro Gly Thr Glu Gly Val Ala
          -10          -5          1

TCC CAA CTC ACC CCA GCC CCA AAA CTC TCC TCT GCT GCT GGC TGG TTA   259
Ser Gln Leu Thr Pro Ala Pro Lys Leu Ser Ser Ala Ala Gly Trp Leu

```

5	10	15	
GAG GTT CCC TTT GAC GCC ATC CCA GCC CCA GGG			292
Glu Val Pro Phe Asp Ala Ile Pro Ala Pro Gly			
20	25		

## (2) INFORMATION FOR SEQ ID NO: 183:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 266 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 96..252
- (C) IDENTIFICATION METHOD: fasta
- (D) OTHER INFORMATION: identity 98  
region 1..157  
id HSU41901  
vrt

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 96..179
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.4  
seq LLRLCLLPGLP/VR

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

AGTGGATSMW AGNNCCGTGT TGGTGAAGCC TCTCCTCGCG AGCAGCGCGC ACCCCTCCAG	60
AGCACCCCGC GGACCCGCAC CTCGGCGTGG CCACC ATG GTC AGG AGA GTT CAG	113
Met Val Arg Arg Val Gln	
-25	
CCG GAT CGG AAA CAG TTG CCA CTG GTC CTA CTG AGA TTG CTC TGC CTT	161
Pro Asp Arg Lys Gln Leu Pro Leu Val Leu Leu Arg Leu Leu Cys Leu	
-20 -15 -10	
CTT CCC ACA GGA CTG CCT GTT CGC AGC GTG GAT TTT AAC CGA GGC ACG	209
Leu Pro Thr Gly Leu Pro Val Arg Ser Val Asp Phe Asn Arg Gly Thr	
-5 1 5 10	
GAC AAC ATC ACC GTG AGG CAG GGG GAC ACA GCC ATC CTC AGA TTT CTC	257
Asp Asn Ile Thr Val Arg Gln Gly Asp Thr Ala Ile Leu Arg Phe Leu	
15 20 25	
VNC TCA GGG	266
Xaa Ser Gly	

## (2) INFORMATION FOR SEQ ID NO: 184:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 332 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 52..327
- (C) IDENTIFICATION METHOD: fasta
- (D) OTHER INFORMATION: identity 97  
region 4..280  
id HUMGPCRB  
vrt

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 293..327
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 1..35  
id T29782  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 180..236
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.4  
seq LVFIIGLVGNLLA/LV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

```

AACTTCNNNN HTGGACAACT ACTCACAGCT ACTACACWRA GACCCGAATM GAGTCACTGA    60
TATACACCTG GACCACCACC AATGGATATA CAAATGGCAA ACAATTTTAC TCCGCCTCTG    120
CAACTCCTCA GGGAAATGAC TGTGACCTCT ATGCACATCA CAGCACGGCC AGGATAGTA    179
ATG CCT CTG CAT TAC AGC CTC GTC TTC ATC ATT GGG CTC GTG GGA AAC    227
Met Pro Leu His Tyr Ser Leu Val Phe Ile Ile Gly Leu Val Gly Asn
      -15                      -10                      -5

TTA CTA GCC TTG GTC GTC ATT GTT CAA AAC AGG AAA AAA ATC AAC TCT    275
Leu Leu Ala Leu Val Val Ile Val Gln Asn Arg Lys Lys Ile Asn Ser
      1                      5                      10

ACC ACC CTC TAT TCA ACA AAT TTG GTT ATT TCT GAT ATA CTT TTT ASC    323
Thr Thr Leu Tyr Ser Thr Asn Leu Val Ile Ser Asp Ile Leu Phe Xaa

```

15

20

25

ACC GTC GGG  
Thr Val Gly  
30

332

## (2) INFORMATION FOR SEQ ID NO: 185:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 273 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 54..272
- (C) IDENTIFICATION METHOD: fasta
- (D) OTHER INFORMATION: identity 97  
region 291..509  
id D82060  
vrt

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 82..243
- (C) IDENTIFICATION METHOD: fasta
- (D) OTHER INFORMATION: identity 98  
region 1..162  
id HUMMHCRING  
vrt

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 61..150
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.5  
seq WATLGLLVAGLGG/HD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

```
AAAAGGCGGG ACTGCCACGT CCAAGCAAAC CGGGAAAGGA GAGGATCCCG GAGCCGCGTG      60
ATG GCC AGA GGC CTG GGG GCC CCC CAC TGG GTG GCC GTG GGA CTG CTG      108
Met Ala Arg Gly Leu Gly Ala Pro His Trp Val Ala Val Gly Leu Leu
-30                      -25                      -20                      -15
ACC TGG GCG ACC TTG GGG CTT CTG GTG GCT GGA CTC GGG GGT CAT GAC      156
Thr Trp Ala Thr Leu Gly Leu Leu Val Ala Gly Leu Gly Gly His Asp
-10                      -5                      1
GAC CTG CAC GAC GAT CTG CAA GAG GAC TTC CAT GGC CAC AGC CAC AGG      204
```

```

Asp Leu His Asp Asp Leu Gln Glu Asp Phe His Gly His Ser His Arg
   5                               10                               15
CAC TCA CAT GAA GAT TTC CAC CAT GGC CAM AGC CAT GCC CAT GGC CAT   252
His Ser His Glu Asp Phe His His Gly Xaa Ser His Ala His Gly His
   20                               25                               30
GGC CAC AMT CAC GAG AGC ATG                                           273
Gly His Xaa His Glu Ser Met
   35                               40

```

## (2) INFORMATION FOR SEQ ID NO: 186:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -14..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 11.8  
seq VLVALILLHSALA/QS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

```

Met Val Leu Val Ala Leu Ile Leu Leu His Ser Ala Leu Ala Gln Ser
      -10                               -5                               1
Arg Arg Asp Phe Ala Pro Pro Gly Gln Gln Lys Arg Glu Ala Pro Gly
   5                               10                               15

```

## (2) INFORMATION FOR SEQ ID NO: 187:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -21..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 10  
seq LLLCLQWPEAAG/KD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

```

Met Ala Gln His His Leu Trp Ile Leu Leu Leu Cys Leu Gln Thr Trp
  -20                      -15                      -10

Pro Glu Ala Ala Gly Lys Asp Ser Glu Ile Phe Thr Val Asn Gly Ile
  -5                      1                      5                      10

Leu Gly Glu Ser Val Thr Phe Pro Val Asn Ile Gln Glu Pro Arg Gln
                      15                      20                      25

Val Lys Ile Ile Ala Trp Thr Ser Lys Thr Ser Val Ala Tyr Val Thr
                      30                      35                      40

Pro Gly Glu Arg
  45

```

(2) INFORMATION FOR SEQ ID NO: 188:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -19..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 10  
seq FLLLVTAPRCILS/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

```

Met Lys Asp Leu Trp Ile Phe Leu Leu Leu Val Thr Ala Pro Arg Cys
  -15                      -10                      -5

Ile Leu Ser Gln Val Gln Leu Gln Glu Ser Gly Pro Arg Leu Val Arg
  1                      5                      10

Pro Ser Glu Thr Val Ser Leu Ser Cys Thr Val Ser Gly Asp Ser Val
  15                      20                      25

Ser Ser Gly Asp His Tyr Trp Thr Trp Leu Arg Gln Pro Pro Gly Gly
  30                      35                      40                      45

Gly Leu Glu Trp Ile Gly Tyr Ile Tyr Thr Thr Gly Lys Ile Asp Tyr
  50                      55                      60

```

Asn Pro Ser Xaa Arg Arg Arg Val Thr Ile Ser Val Asp Thr Ser Lys  
65 70 75  
Asn Leu Phe Ser Leu Thr Arg  
80

## (2) INFORMATION FOR SEQ ID NO: 189:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 79 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Placenta

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -21..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10  
seq LLLCLQTWPEAAG/KD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

Met Ala Gln His His Leu Trp Ile Leu Leu Leu Cys Leu Gln Thr Trp  
-20 -15 -10  
Pro Glu Ala Ala Gly Lys Asp Ser Glu Ile Phe Thr Val Asn Gly Ile  
-5 1 5 10  
Leu Gly Glu Ser Val Thr Phe Pro Val Asn Ile Gln Glu Pro Arg Gln  
15 20 25  
Val Lys Ile Ile Ala Trp Thr Ser Lys Thr Ser Val Ala Tyr Val Thr  
30 35 40  
Pro Gly Asp Ser Glu Thr Ala Pro Val Val Thr Val Thr His Met  
45 50 55

## (2) INFORMATION FOR SEQ ID NO: 190:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.9  
seq FLFVVAAATGVQS/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

```

Met Asp Trp Thr Trp Arg Phe Leu Phe Val Val Ala Ala Ala Thr Gly
      -15                      -10                      -5

Val Gln Ser Gln Val Gln Leu Val Gln Ser Gly Ala Glu Val Val Lys
      1                      5                      10

Pro Gly Ser Ser Val Lys Val Ser Cys Lys Thr Ser Gly Asp Gly Phe
      15                      20                      25

Ser Lys Tyr Pro Ile Asn Trp Val Gln Gly
      30                      35

```

## (2) INFORMATION FOR SEQ ID NO: 191:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.6  
seq GLLLLCLLPURLA/LV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

```

Met Ser Ile Cys Phe Leu Gly Leu Leu Leu Cys Leu Leu Pro His
      -15                      -10                      -5

Arg Leu Ala Leu Val Gln Lys His Ser Ser Pro Ser Ser Arg Leu Leu
      1                      5                      10

Leu Ile Pro Val Val Gln Cys Leu Leu Ala Leu Glu Phe Leu Gln Asp
      15                      20                      25

Pro Tyr Leu Asp Ile Phe Asn Leu Pro Leu Pro Pro Pro Trp
      30                      35                      40

```



## (2) INFORMATION FOR SEQ ID NO: 192:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -17..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.2  
seq LVLLILPLLSSLS/KV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

Met Ile Gly Phe Leu Val Leu Leu Ile Leu Pro Leu Leu Ser Ser Leu  
-15 -10 -5

Ser Lys Val Ser Ser Lys  
1 5

## (2) INFORMATION FOR SEQ ID NO: 193:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -31..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.9  
seq LLMSLLVSTVTWQ/IS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

Met Gln Cys Leu Leu Ser Val Leu Met Ala Gln Phe Ile Xaa His Phe  
-30 -25 -20

Leu Ser Leu Leu Met Ser Leu Leu Val Ser Thr Val Thr Trp Gln Ile  
-15 -10 -5 1

Ser Arg Thr Pro Trp His Gly

5

## (2) INFORMATION FOR SEQ ID NO: 194:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 87 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.9  
seq WIFFLATLKGVC/QV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

Met Glu Leu Gly Leu Ser Trp Ile Phe Phe Leu Ala Thr Leu Lys Gly  
                  -15                  -10                  -5

Val Gln Cys Gln Val Arg Leu Leu Glu Ser Ala Gly Gly Leu Gln Glu  
                  1                  5                  10

Pro Gly Gly Ala Leu Arg Leu Ser Cys Ala Val Ser Gly Phe Ile Phe  
          15                  20                  25

Asn Asp Phe Ala Met His Trp Val Arg Gln Thr Pro Gly Lys Gly Leu  
          30                  35                  40                  45

Glu Trp Val Ala Gly Ile Asn Trp Asp Gly Xaa Ile Leu Gly Tyr Ala  
                  50                  55                  60

Asp Ser Val Lys Gly Arg Arg  
                  65

## (2) INFORMATION FOR SEQ ID NO: 195:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -15..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.6  
seq SVSLALLSGWVGS/RQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

```

Met Val Ser Val Ser Leu Ala Leu Leu Ser Gly Trp Val Gly Ser Arg
-15                -10                -5                1

Gln Gly Arg Val Gly Leu Ser Thr Leu Val Thr Leu Gly Leu Val Ser
          5                10                15

Trp Cys Trp Arg Met Val Arg Thr Gln Ala Leu Glu Gly Phe Leu Ser
      20                25                30

Val Lys Tyr Tyr Ser Ala Phe Ser Ala Asp Gln
      35                40

```

(2) INFORMATION FOR SEQ ID NO: 196:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Placenta

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -21..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.5  
seq LPLLLSWVAGGFG/NA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

```

Met Pro Leu Pro Trp Ser Leu Ala Leu Pro Leu Leu Ser Trp Val
-20                -15                -10

Ala Gly Gly Phe Gly Asn Ala Ala Ser Ala Arg His His Gly Leu Leu
-5                1                5                10

Ala Ser Ala Arg Gln Pro Gly Val Cys His Tyr Gly Thr Lys Leu Ala
      15                20                25

Cys Cys Tyr Gly Trp Arg Arg Asn Ser
      30                35

```

(2) INFORMATION FOR SEQ ID NO: 197:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 114 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -80..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.4  
seq FVVFSFLICAMA/GD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

Met Val Ser Asn Phe Phe His Val Ile Gln Val Phe Glu Lys Ser Ala  
-80 -75 -70 -65

Thr Leu Ile Ser Lys Thr Glu His Ile Gly Phe Val Ile Tyr Ser Trp  
-60 -55 -50

Xaa Lys Ser Thr Thr His Leu Gly Ser Arg Arg Lys Phe Ala Ile Ser  
-45 -40 -35

Ile Tyr Leu Ser Glu Val Ser Leu Gln Lys Tyr Asp Cys Pro Phe Ser  
-30 -25 -20

Gly Thr Ser Phe Val Val Phe Ser Leu Phe Leu Ile Cys Ala Met Ala  
-15 -10 -5

Gly Asp Val Val Tyr Ala Asp Ile Lys Thr Val Arg Thr Ser Pro Leu  
1 5 10 15

Glu Leu Ala Xaa Xaa Leu Gln Arg Ser Xaa Xaa Phe Asn Phe Ser Xaa  
20 25 30

Xaa Arg

## (2) INFORMATION FOR SEQ ID NO: 198:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: -44..-1  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 8.2  
 seq ICLACVLFPLLRT/SD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

```

Met Arg Xaa Phe Trp Phe Leu Met Tyr Pro Phe Arg Phe His Asp Cys
      -40                -35                -30

Lys Gln Lys Tyr Asp Leu Tyr Ile Ser Ile Ala Gly Trp Leu Ile Ile
      -25                -20                -15

Cys Leu Ala Cys Val Leu Phe Pro Leu Leu Arg Thr Ser Asp Asp Thr
      -10                -5                1

Pro Gly Asn Arg Thr Lys Cys Phe Val Asp Leu Pro Thr Arg Asn Val
      5                10                15                20

Asn Leu Ala Gln Ser Val Val Met Met Thr Ile Gly Glu Leu Ile Gly
      25                30                35
  
```

(2) INFORMATION FOR SEQ ID NO: 199:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 amino acids  
 (B) TYPE: AMINO ACID  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: -17..-1  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 7  
 seq CCLFTCFIPICIS/CK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

```

Met Val Ser Leu Cys Cys Leu Phe Thr Cys Phe Phe Ile Pro Cys Ile
      -15                -10                -5

Ser Cys Lys Leu Glu Met Trp Gly Leu Asp Glu Pro Lys Val Lys Pro
      1                5                10                15

Phe Trp Gln Glu Cys Val Leu Gly Asp Val Val Gly Xaa Ile Leu Gln
      20                25                30

His Arg Arg Gln Pro Pro Val Pro Arg Ser Ile Leu Val Met Gly Ala
      35                40                45
  
```

## (2) INFORMATION FOR SEQ ID NO: 200:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Placenta

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -27..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.9  
seq ILLLVITYSPIAYS/HS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

Met Asp Phe Phe Phe Leu Glu Arg Ser Tyr Trp Gly Lys Met Ile Leu  
-25 -20 -15

Leu Leu Val Thr Tyr Ser Pro Ile Ala Tyr Ser His Ser Arg  
-10 -5 1

## (2) INFORMATION FOR SEQ ID NO: 201:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -26..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.9  
seq LWVLLLCARVVTL/LV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

Met Thr Met Arg His Asn Trp Thr Pro Asp Leu Ser Pro Leu Trp Val  
-25 -20 -15

Leu Leu Leu Cys Ala His Val Val Thr Leu Leu Val Arg Ala Thr Pro  
-10 -5 1 5

Val Ser Gln Thr Xaa Thr Ala Ala Thr Ala Ser Val Arg Ser Thr Lys  
                   10                                  15                                  20

Asp Pro Cys Pro Thr Gln Gly  
                   25

## (2) INFORMATION FOR SEQ ID NO: 202:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -48..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.8  
seq ILRMLLSLQPVLQ/DA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

Met Asp Asn Met Ser Gly Gly Lys Val Asp Glu Ala Leu Val Lys Ser  
                   -45                                  -40                                  -35

Ser Cys Leu His Pro Trp Ser Lys Arg Asn Asp Val Ser Met Gln Cys  
                   -30                                  -25                                  -20

Ser Gln Asp Ile Leu Arg Met Leu Leu Ser Leu Gln Pro Val Leu Gln  
                   -15                                  -10                                  -5

Asp Ala Ile Gln Lys Lys Arg Thr Val Arg Gln  
   1                                  5                                  10

## (2) INFORMATION FOR SEQ ID NO: 203:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide

(B) LOCATION: -45..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.7  
seq AALVLWTLPGAQR/RG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

Met Xaa Leu Gln Gly Gln Glu Ala Thr Gly Lys Val Leu Ile Lys Ile  
-45                      -40                      -35                      -30  
His Lys Asp Thr Ser Gln Val Pro Thr Ala Xaa Gly Asp Ala Ser Ile  
                    -25                      -20                      -15  
Ala Ala Leu Val Leu Trp Thr Leu Pro Gly Ala Gln Arg Arg Gly Glu  
                    -10                      -5                      1  
Phe Ala Pro Lys Gly Ala Pro Met Thr Asn Arg  
                    5                      10

(2) INFORMATION FOR SEQ ID NO: 204:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -25..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.6  
seq ILVLILFPTSCVM/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

Met Thr Glu His Ser Leu Thr His Gln Gly Ile Pro Ile Leu Val Leu  
-25                      -20                      -15                      -10  
Ile Leu Phe Pro Thr Ser Cys Val Met Gln Val Leu Trp  
                    -5                      1

(2) INFORMATION FOR SEQ ID NO: 205:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN



## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.5  
seq RFIFLTSLQLISS/SY

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

```

Met Tyr Ile Gly Gly Leu Arg Phe Ile Phe Leu Thr Ser Leu Gln Leu
      -15                      -10                      -5
Ile Ser Ser Ser Tyr Val Thr Thr Leu Leu Lys Lys Asn Thr Leu Arg
      1                      5                      10

```

## (2) INFORMATION FOR SEQ ID NO: 206:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -38..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.5  
seq LIFFSLIFLNLFA/IS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

```

Met Ser Val Ser Leu Lys His Ile His Leu His Phe Ile Ile Met Ser
      -35                      -30                      -25
Val Leu Val Phe Trp Asn Cys Ser His Leu Ile Phe Phe Ser Leu Ile
      -20                      -15                      -10
Phe Leu Asn Leu Phe Ala Ile Ser Trp
      -5                      1

```

## (2) INFORMATION FOR SEQ ID NO: 207:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids

(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -22..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.3  
seq MVSFLSXPFLCSA/KP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

Met Xaa Xaa Leu Gly Xaa Xaa Arg Phe Met Val Ser Phe Leu Ser Xaa  
-20 -15 -10

Pro Phe Leu Cys Ser Ala Lys Pro Ser Ser Gly  
-5 1 5

(2) INFORMATION FOR SEQ ID NO: 208:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 71 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -19..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.3  
seq ILVSVAAATGAHS/QL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

Met Asp Trp Thr Trp Tyr Ile Leu Val Ser Val Ala Ala Ala Thr Gly  
-15 -10 -5

Ala His Ser Gln Leu Gln Leu Leu Gln Ser Gly Ser Asp Ile Lys Lys  
1 5 10

Pro Gly Ala Ser Met Asn Val Ser Cys Lys Ala Ser Gly Gly Ser Ile  
15 20 25

Ser Thr Arg Gly Ile Ser Trp Val Arg Gln Val Pro Gly Gln Gly Leu  
30 35 40 45

Glu Trp Met Gly Trp Ile Gly  
50

(2) INFORMATION FOR SEQ ID NO: 209:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -45..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.3  
seq VACVLSSLIIVNS/AH

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

Met	Ile	Ser	Lys	Phe	Ser	Ser	Lys	Ala	Tyr	Ser	Val	Arg	Gly	Leu	Glu
-45					-40					-35					-30
Leu	Phe	Ser	Leu	Leu	Pro	Ile	Asn	Pro	Ser	Pro	Asn	Ser	Ala	Ile	Xaa
			-25					-20						-15	
Val	Ala	Cys	Val	Leu	Ser	Ser	Leu	Ile	Ala	Val	Asn	Ser	Ala	His	Pro
		-10					-5							1	
Glu	Ser	Thr	Ile	Asp	Thr	Arg	Trp								
5					10										

(2) INFORMATION FOR SEQ ID NO: 210:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -28..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.3

seq LLFLIFSLNLNRG/VG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

Met Val Leu Leu Gly Ala Phe Gly Ser Cys Ile Lys Ser Phe Ser Leu  
                   -25                  -20                  -15

Leu Phe Leu Ile Phe Ser Leu Asn Leu Asn Arg Gly Val Gly  
                   -10                  -5                  1

(2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 105 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -35..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.3  
seq ALKLLLSPGXSGS/SS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

Met Ala Ala Arg Gln Ala Val Gly Ser Gly Ala Gln Glu Thr Cys Gly  
   -35                  -30                  -25                  -20

Leu Asp Arg Ile Leu Glu Ala Leu Lys Leu Leu Leu Ser Pro Gly Xaa  
                   -15                  -10                  -5

Ser Gly Ser Ser Ser Leu Gln Val Thr Lys His Asp Val Leu Leu Ala  
                   1                  5                  10

Thr Leu Lys Ser Asn Leu Ser Ala Leu Glu Asp Lys Phe Leu Lys Asp  
   15                  20                  25

Pro Gln Trp Lys Asn Leu Lys Leu Leu Arg Asp Glu Ile Ala Asp Lys  
   30                  35                  40                  45

Ala Glu Trp Pro Gln Asn Ser Val Asp Val Thr Trp Ser Phe Thr Ser  
                   50                  55                  60

Gln Thr Leu Leu Leu Leu Cys Leu  
                   65                  70

(2) INFORMATION FOR SEQ ID NO: 212:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.2  
seq LALFLMALGFSCI/HK

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

Met Ser Thr Gln Lys Gly Leu Ala Leu Phe Leu Met Ala Leu Gly Phe  
                  -15                  -10                  -5

Ser Cys Ile His Lys Lys Phe Gln Glu Ser Glu Glu Gly Lys His His  
                  1                          5                          10

Met Gly Gly Ile Asn Arg Ser His Trp Val Lys Ser Arg Lys Ser Cys  
          15                          20                          25

Leu Ile Asn Ser Gln Arg Lys  
          30                          35

## (2) INFORMATION FOR SEQ ID NO: 213:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -25..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.1  
seq YFLIVFFVFLCNC/HQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

Met Lys Asp Val Glu Ile Ile Met Ile Phe His Gly Tyr Phe Leu Ile  
-25                          -20                          -15                          -10

Val Phe Phe Val Phe Leu Cys Asn Cys His Gln  
-5 1

## (2) INFORMATION FOR SEQ ID NO: 214:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -42..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.1  
seq AILLLQSQCAYWA/LP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

Met Cys Phe Pro Glu His Arg Arg Gln Met Tyr Ile Gln Asp Arg Leu  
-40 -35 -30

Asp Ser Val Thr Arg Arg Ala Arg Gln Gly Arg Ile Cys Ala Ile Leu  
-25 -20 -15

Leu Leu Gln Ser Gln Cys Ala Tyr Trp Ala Leu Pro  
-10 -5 1

## (2) INFORMATION FOR SEQ ID NO: 215:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -23..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.1  
seq SSILSTFVSWLSA/FY

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

Met Leu Val Val Lys Gln Cys Phe Ser Asp Ser Ser Ile Leu Ser Thr  
                   -20                                  -15                                  -10

Phe Val Ser Trp Leu Ser Ala Phe Tyr Cys Lys Glu Gly Pro Ser Ser  
           -5                                  1                                  5

Gly  
 10

(2) INFORMATION FOR SEQ ID NO: 216:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 104 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -32..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.1  
 seq LPLLTSALHGLQQ/QH

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

Met Ile Xaa Leu Arg Asp Thr Ala Ala Ser Leu Arg Leu Glu Arg Asp  
           -30                                  -25                                  -20

Thr Arg Gln Leu Pro Leu Leu Thr Ser Ala Leu His Gly Leu Gln Gln  
           -15                                  -10                                  -5

Gln His Pro Ala Phe Ser Gly Val Ala Arg Leu Ala Lys Arg Trp Val  
       1                                  5                                  10                                  15

Arg Ala Gln Leu Leu Gly Glu Gly Phe Ala Asp Glu Ser Leu Asp Leu  
                   20                                  25                                  30

Val Ala Ala Ala Leu Phe Leu His Pro Glu Pro Phe Thr Pro Pro Ser  
           35                                  40                                  45

Ser Pro Gln Val Gly Phe Leu Arg Phe Leu Phe Leu Val Ser Thr Phe  
           50                                  55                                  60

Asp Trp Lys Asn Asn Pro Leu Gly  
       65                                  70

(2) INFORMATION FOR SEQ ID NO: 217:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -14..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9  
seq ITMMLALISVCLF/AF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

Met Ile Thr Met Met Leu Ala Leu Ile Ser Val Cys Leu Phe Ala Phe  
                  -10                  -5                  1

Trp

(2) INFORMATION FOR SEQ ID NO: 218:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -15..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9  
seq LLTLVQCSDLCP/CS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

Met Trp Leu Leu Thr Leu Val Gln Cys Ser Asp Leu Cys Pro Ser Cys  
-15                  -10                  -5                  1

Ser Gln Ala Leu Thr Leu Val Leu Val Ser Phe Ser Glu Val Arg Asp  
                  5                  10                  15

Leu Ala Glu Thr Ser Leu Ser Ser Asn Leu Lys Asn Ser Leu Phe Ile  
                  20                  25                  30

Val Leu Lys Arg  
                  35



## (2) INFORMATION FOR SEQ ID NO: 219:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -31..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9  
seq LLFACLTMLLVKT/CQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

Met Arg Val His Leu Phe Pro Tyr Leu Cys Gln Pro Ser Val Leu Ser  
-30 -25 -20

Asn Phe Leu Leu Phe Ala Cys Leu Thr Met Leu Leu Val Lys Thr Cys  
-15 -10 -5 1

Gln Glu Ser Pro Lys Ser Pro Leu Ser Leu Met Ile Cys Gln Thr Tyr  
5 10 15

Arg Ile Gly  
20

## (2) INFORMATION FOR SEQ ID NO: 220:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -15..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.7  
seq PLCFLILPYPVLS/SH

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

```

Met Ile Pro Leu Cys Phe Leu Ile Leu Pro Tyr Pro Val Leu Ser Ser
-15          -10          -5          1

His Asp His Asn Ser Leu Gly Leu Leu Ala Asp Lys Val Ala Asn Glu
      5          10          15

Ile Asn Arg Ser Asn Cys Arg Val Tyr Ala His Ser His Ser Gly
    20          25          30

```

## (2) INFORMATION FOR SEQ ID NO: 221:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -32..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.5  
seq CLLSXPSTRKSQA/CM

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

```

Met Ala Gly Ser Arg Leu Pro Arg Gln Leu Phe Leu Gln Gly Val Xaa
    -30          -25          -20

Ala Ser Ser Cys Leu Leu Ser Xaa Pro Ser Thr Arg Lys Ser Gln Ala
-15          -10          -5

Cys Met Ala Pro Arg Ala Trp
  1          5

```

## (2) INFORMATION FOR SEQ ID NO: 222:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -28..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.4  
seq FVLHLLAQLVCC/FY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

```

Met Tyr Ile Cys Phe Cys Leu Glu Ser Phe Glu Ile Lys Cys Gly Phe
  -25                      -20                      -15

Val Leu His Leu Leu Ala Gln Asp Leu Val Cys Cys Phe Tyr Leu Arg
  -10                      -5                      1

Thr Xaa Xaa
  5

```

(2) INFORMATION FOR SEQ ID NO: 223:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -20..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.4  
seq LNAFTLLVWLSLS/KN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

```

Met His Phe Ile Leu His Asn Leu Asn Ala Phe Thr Leu Leu Val Trp
-20                      -15                      -10                      -5

Leu Ser Leu Ser Lys Asn Thr Val Pro Arg Pro Ala Val Leu Ala Ser
      1                      5                      10

Ala Ala Trp
      15

```

(2) INFORMATION FOR SEQ ID NO: 224:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -35..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 5.2  
seq IAPLFTLLPKSIP/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

Met Ser Phe Phe Pro Phe Asn Arg Ser Leu Asn Ser Asn Pro His Pro  
-35 -30 -25 -20

Asn Leu Leu Phe Pro Asn Ile Ala Pro Leu Phe Thr Leu Leu Pro Lys  
-15 -10 -5

Ser Ile Pro Ala Pro  
1

(2) INFORMATION FOR SEQ ID NO: 225:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 54 amino acids

(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -23..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 5.2  
seq LLDLHCFCSLAKT/KN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

Met Val Val Trp Val Leu Glu Val Arg Phe Leu Leu Asp Leu His Cys  
-20 -15 -10

Phe Cys Ser Leu Ala Lys Thr Lys Asn Gly Leu Ser Trp Gly Leu Pro  
-5 1 5

Gln Lys Val Ala Leu Cys Thr Pro Cys Ser Ala Pro Ala Leu Phe Trp  
10 15 20 25

Phe Gly Phe His Ile Leu  
30

## (2) INFORMATION FOR SEQ ID NO: 226:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -24..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.2  
seq PGLCCPALGSAWS/KN

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

Met Val Cys Gly Trp Trp Thr Gln Gly Pro Val Pro Gly Leu Cys Cys  
                  -20                  -15                  -10

Pro Ala Leu Gly Ser Ala Trp Ser Lys Asn Lys Ser Xaa Pro Val Pro  
                  -5                                  1                                  5

Cys Cys Gly Pro Tyr Met Val Ala Asn Leu Gly  
          10                                  15

## (2) INFORMATION FOR SEQ ID NO: 227:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -26..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.2  
seq FECALVSASLTTA/GT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

Met Gly Arg Ala Phe Pro Ser Arg His Lys Thr Ala Arg Phe Glu Cys

-25                      -20                      -15  
Ala Leu Val Ser Ala Ser Leu Thr Thr Ala Gly Thr Pro Gly Lys Asn  
-10                      -5                      1                      5  
Leu Xaa Ser Tyr Asn Ser Ala Glu Ala Arg His Ile  
10                      15

## (2) INFORMATION FOR SEQ ID NO: 228:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Placenta

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -18..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5  
seq LCXXLLCVLFVSH/FY

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

Met Gly Leu Lys Ala Leu Cys Xaa Xaa Leu Leu Cys Val Leu Phe Val  
-15                      -10                      -5  
Ser His Phe Tyr Thr Pro Thr  
1                      5

## (2) INFORMATION FOR SEQ ID NO: 229:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 94 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -76..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5  
seq FPLLALLFEKCEQ/ST

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

```

Met Met Ala Thr Gln Thr Leu Ser Ile Asp Ser Tyr Gln Asp Gly Gln
-75                      -70                      -65

Gln Met Gln Val Val Thr Glu Leu Lys Thr Glu Gln Asp Pro Asn Cys
-60                      -55                      -50                      -45

Ser Glu Pro Asp Ala Glu Gly Val Ser Pro Pro Pro Val Glu Ser Gln
                      -40                      -35                      -30

Thr Pro Met Asp Val Asp Lys Gln Ala Ile Tyr Arg His Pro Leu Phe
                      -25                      -20                      -15

Pro Leu Leu Ala Leu Leu Phe Glu Lys Cys Glu Gln Ser Thr Gln Gly
                      -10                      -5                      1

Ser Glu Gly Thr Thr Ser Ala Ser Phe Asp Val Asp Ile Gly
5                      10                      15

```

(2) INFORMATION FOR SEQ ID NO: 230:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -21..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.9  
seq SVFLSGSVCLSFL/SE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

```

Met Ser Pro Ser Gln Leu Thr Cys Ser Val Phe Leu Ser Gly Ser Val
-20                      -15                      -10

Cys Leu Ser Phe Leu Ser Glu His Arg Thr Tyr Phe Phe Cys Pro Leu
-5                      1                      5                      10

```

(2) INFORMATION FOR SEQ ID NO: 231:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia
- (ix) FEATURE:  
(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -30..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.8  
seq FCSLLCLRTQLFP/HG
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

Met Leu Gln Ala Leu Ala Pro Ala His His Leu Cys Ser Leu Lys Arg  
-30 -25 -20 -15

Ser Phe Cys Ser Leu Leu Cys Leu Arg Thr Gln Leu Phe Pro His Gly  
-10 -5 1

(2) INFORMATION FOR SEQ ID NO: 232:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 35 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia
- (ix) FEATURE:  
(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -14..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.8  
seq LFLKYLWRS�CRG/GI
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

Met Leu Phe Leu Lys Tyr Leu Trp Arg Ser Leu Cys Arg Gly Gly Ile  
-10 -5 1

Ile Arg Met Asn His Pro Gly Cys Ser Gln Arg Ile Arg Asp Ser Leu  
5 10 15

Cys Asp Leu  
20

(2) INFORMATION FOR SEQ ID NO: 233:

- (i) SEQUENCE CHARACTERISTICS:



- (A) LENGTH: 95 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -37..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.8  
seq AILIRPLVSVSGS/GP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

```
Met Ala Leu Leu Ala Met His Ser Trp Arg Trp Ala Ala Ala Ala Ala
   -35                      -30                      -25

Ala Phe Glu Lys Arg Arg His Ser Ala Ile Leu Ile Arg Pro Leu Val
   -20                      -15                      -10

Ser Val Ser Gly Ser Gly Pro Gln Trp Arg Pro His Gln Leu Gly Ala
   -5                      1                      5                      10

Leu Gly Thr Ala Arg Ala Tyr Gln Ile Pro Glu Ser Leu Lys Ser Ile
      15                      20                      25

Thr Trp Gln Arg Leu Gly Lys Gly Asn Ser Gly Gln Phe Leu Asp Ala
      30                      35                      40

Ala Lys Ala Leu Gln Val Trp Pro Leu Ile Glu Lys Arg Thr Trp
   45                      50                      55
```

(2) INFORMATION FOR SEQ ID NO: 234:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Placenta

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -38..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.3  
seq LASLFGLDQXAAG/HG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:



- (B) LOCATION: -15..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.7  
seq IFHVLIHSSSFS/CE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

```

Met Thr Ile Phe His Val Leu Ile Ala His Ser Ser Ser Phe Ser Cys
-15                -10                -5                1
Glu Val Ile Val Lys Val Phe Cys Pro Phe Leu Gly Cys Leu Ser Phe
          5                10                15
His Tyr Leu Tyr Ser Leu Leu Glu Phe Phe Ile Leu Asn Thr Ser Pro
      20                25                30
Ser Met
      35

```

(2) INFORMATION FOR SEQ ID NO: 237:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -15..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.7  
seq WQLXGFCGSYSA/AQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

```

Met His Trp Gln Leu Leu Xaa Gly Phe Cys Gly Ser Tyr Ser Ala Ala
-15                -10                -5                1
Gln Ala Glu Ala Gln Thr Leu Pro Gly Leu His Ser Lys Tyr Asn Thr
          5                10                15
His Gly

```

(2) INFORMATION FOR SEQ ID NO: 238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 65 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -37..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 4.6  
seq FVLLFFFSXLXY/FM

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

```
Met Thr Met Met Val Met Ala Ser Phe Leu Pro Arg Asn Thr Met Tyr
      -35                -30                -25

Thr Asn Thr Met Asn Tyr Ser Ile Phe Val Phe Leu Leu Phe Phe Phe
      -20                -15                -10

Ser Xaa Leu Xaa Tyr Phe Met Tyr Lys Thr Ser His Phe Ser Pro Ser
      -5                  1                  5                  10

Xaa Ile Cys Tyr Phe Ser Pro Met Xaa Xaa Xaa Xaa Asp Leu Pro Asn
      15                  20                  25

Gly
```

(2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 100 amino acids

(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -61..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 4.5  
seq LVTRLALCQSPRA/GQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

```
Met Pro Ser Gln Thr Leu Ser Gln Pro Arg Ile Ser Val Leu His Gly
      -60                -55                -50

Asp Leu Val Pro Ala Gly Met Ala Val Gln Glu Ile Gly Ala Gln Met
      -45                -40                -35                -30
```

Val Leu Pro Cys Glu Val Val Ser Gly Ser Gly Leu Thr Arg Glu His  
                   -25                  -20                  -15

Leu Val Thr Arg Leu Ala Leu Cys Gln Ser Pro Arg Ala Gly Gln His  
                   -10                  -5                  1

Gly Ala Asp Ser Glu Glu Glu Ala Phe Gly Ile Leu Pro Val Arg His  
           5                  10                  15

Ser His Arg Leu Ser Ala Cys His Thr Pro Gly Glu Leu Arg Phe Ser  
   20                  25                  30                  35

Glu Trp Thr Cys

## (2) INFORMATION FOR SEQ ID NO: 240:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -38..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.5  
seq LLMITVTVGPGAS/GV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

Met Ser Leu Arg Val His Thr Leu Pro Thr Leu Leu Gly Ala Val Val  
           -35                  -30                  -25

Arg Pro Gly Cys Arg Glu Leu Leu Cys Leu Leu Met Ile Thr Val Thr  
       -20                  -15                  -10

Val Gly Pro Gly Ala Ser Gly Val Cys Pro Ser Gly  
   -5                  1                  5

## (2) INFORMATION FOR SEQ ID NO: 241:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -18..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.5  
seq SLLLLGRWLTLS/SG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

Met Ile Tyr Leu Thr Ser Leu Leu Leu Gly Arg Trp Leu Thr Leu  
-15 -10 -5

Thr Ser Ser Gly  
1

(2) INFORMATION FOR SEQ ID NO: 242:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -22..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.4  
seq GFLLCPLVCGLRR/WT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

Met Asn Trp Asn Val Arg Gly Thr Arg Gly Phe Leu Leu Cys Pro Leu  
-20 -15 -10

Val Cys Gly Leu Arg Arg Trp Thr  
-5 1

(2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -42..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.4  
seq LVCLTFITATTHE/QP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

```
Met Glu Gln Ala Ala Leu Glu Val Val Ser Pro Leu Pro Arg Arg Cys
   -40                      -35                      -30

Ser Val Arg Ser Pro Val Thr Thr Cys Cys Ala Lys Asp Leu Val Cys
   -25                      -20                      -15

Leu Thr Phe Ile Thr Ala Thr Thr His Glu Gln Pro
   -10                      -5                      1
```

## (2) INFORMATION FOR SEQ ID NO: 244:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 100 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -84..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.4  
seq GLVQLHATXLALG/KV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

```
Met Ile Ile Pro Leu Pro Ser Leu Val Gly Cys Trp Glu Gly Gly Asn
   -80                      -75                      -70

Gly Lys Gly Leu Met Val Ser Asp Thr Thr Cys Trp Thr Leu Ala Ser
   -65                      -60                      -55

Ser Asn Val Pro Ser Pro Ser Pro Ala Pro Thr Leu Gly Arg Gly Ala
   -50                      -45                      -40

Pro Ser His Thr Pro Gln Lys Lys Pro Thr Ile Pro Gly Ala Arg His
   -35                      -30                      -25

Arg Pro Ile Ile Leu Pro Lys Gly Leu Val Gln Leu His Ala Thr Xaa
```

-20                      -15                      -10                      -5

Leu Ala Leu Gly Lys Val Cys Leu Pro His Val Pro His His Ala Ser  
                              1                                 5    10    .

Leu Arg Pro Ala  
                      15

(2) INFORMATION FOR SEQ ID NO: 245:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymphocytes

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: -42..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.3  
seq IQTVHIALPGSLG/HP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

Met	Ser	Met	Arg	Leu	Ser	Gly	Glu	Arg	Ile	Tyr	Leu	Leu	Leu	Glu	Val
		-40					-35					-30			
Trp	Leu	Pro	Xaa	Leu	Asn	Phe	Glu	Ser	Val	Leu	His	Phe	Ile	Gln	Thr
	-25					-20					-15				
Val	His	Ile	Ala	Leu	Pro	Gly	Ser	Leu	Gly	His	Pro	Met	Gly	Pro	Cys
-10					-5					1				5	
Ala	Cys	Arg	Pro	Ser	Leu	Ala	His	Pro							
			10				15								

(2) INFORMATION FOR SEQ ID NO: 246:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:



(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: -16..-1  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 4.3  
 seq LLLFCFMPVVINP/DR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

Met Gly Thr Leu Leu Leu Phe Cys Phe Met Pro Val Val Ile Asn Pro  
 -15 -10 -5

Asp Arg  
 1

(2) INFORMATION FOR SEQ ID NO: 247:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acids  
 (B) TYPE: AMINO ACID  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: -22..-1  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 4.1  
 seq GIYLQLFFLSIVS/QP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

Met Val Val Leu Asn Pro Met Thr Leu Gly Ile Tyr Leu Gln Leu Phe  
 -20 -15 -10

Phe Leu Ser Ile Val Ser Gln Pro Thr Phe Ile Asn Ser Val Leu Pro  
 -5 1 5 10

Ile Ser Ala Ala Leu Pro Ser Leu Asp Gln Lys Lys Arg Gly Gly His  
 15 20 25

Lys Ala Cys Cys Leu Leu Thr Pro  
 30

(2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 amino acids  
 (B) TYPE: AMINO ACID  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -36..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 4.1  
seq HWLFLASLSGIKT/YQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

Met Ala Pro His Thr Ala Ser Phe Gly Val Cys Pro Leu Leu Ser Val  
-35 -30 -25

Thr Arg Val Val Ala Thr Glu His Trp Leu Phe Leu Ala Ser Leu Ser  
-20 -15 -10 -5

Gly Ile Lys Thr Tyr Gln Ser Tyr Ile Ser Val Phe Cys Lys Val Thr  
1 5 10

Gly

(2) INFORMATION FOR SEQ ID NO: 249:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -20..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 4.1  
seq SLPCLSFCTLCLV/TP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

Met Ser Tyr Lys Trp Met Pro Ser Leu Pro Cys Leu Ser Phe Cys Thr  
-20 -15 -10 -5

Leu Cys Leu Val Thr Pro Gly  
1

(2) INFORMATION FOR SEQ ID NO: 250:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 25 amino acids  
    (B) TYPE: AMINO ACID  
    (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:  
    (A) ORGANISM: Homo Sapiens  
    (F) TISSUE TYPE: Umbilical cord
- (ix) FEATURE:  
    (A) NAME/KEY: sig\_peptide  
    (B) LOCATION: -22..-1  
    (C) IDENTIFICATION METHOD: Von Heijne matrix  
    (D) OTHER INFORMATION: score 4.1  
                            seq LAGFLLVLYVCLP/HA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

Met Pro Leu Pro Thr Trp Ala Pro Thr Leu Ala Gly Phe Leu Leu Val  
    -20                    -15                    -10

Leu Tyr Val Cys Leu Pro His Ala Gly  
    -5                            1

(2) INFORMATION FOR SEQ ID NO: 251:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 33 amino acids  
    (B) TYPE: AMINO ACID  
    (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:  
    (A) ORGANISM: Homo Sapiens  
    (F) TISSUE TYPE: Umbilical cord
- (ix) FEATURE:  
    (A) NAME/KEY: sig\_peptide  
    (B) LOCATION: -17..-1  
    (C) IDENTIFICATION METHOD: Von Heijne matrix  
    (D) OTHER INFORMATION: score 4.1  
                            seq LLDWIGLKALIRG/HD
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

Met Asn Leu Tyr Leu Leu Asp Trp Ile Gly Leu Lys Ala Leu Ile Arg  
    -15                    -10                    -5

Gly His Asp Ile Lys Ile Gln Ser Leu Cys Pro Ser Pro Cys Leu Pro  
    1                    5                    10                    15

Arg

## (2) INFORMATION FOR SEQ ID NO: 252:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -54..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.1  
seq LLLFCFMPVVINP/DX

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

Met Ser Cys Xaa Val Xaa Asp Ala Xaa Xaa Arg Trp Trp Ala His Xaa  
                  -50                                  -45                                  -40

Leu Ile Ile Gly Trp Xaa His Leu Thr Gln Lys Val His Pro Ile Ala  
                  -35                                  -30                                  -25

Leu Ser His Cys Val Asn Met Gly Thr Leu Leu Leu Phe Cys Phe Met  
                  -20                                  -15                                  -10

Pro Val Val Ile Asn Pro Asp Xaa Gly  
                  -5  1

## (2) INFORMATION FOR SEQ ID NO: 253:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -22..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4  
seq ILAFQTFLNLRA/HL

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

Met Val Pro Asn Leu Cys Gly Arg Gln Ile Leu Ala Phe Gln Thr Phe  
           -20                              -15                              -10

Leu Leu Asn Leu Arg Ala His Leu Phe Gln Leu Ala Ser Arg  
       -5                                  1                                  5

## (2) INFORMATION FOR SEQ ID NO: 254:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -14..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4  
seq FSLIIFFFPPSSP/XA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

Met Phe Ser Leu Ile Ile Phe Phe Phe Pro Pro Ser Ser Pro Xaa Ala  
                               -10                              -5                              1

Asn Pro Phe Pro Ser Tyr Leu Gln Asn Ile Leu Tyr Leu Lys Phe Val  
                   5                              10                              15

His Xaa Ser His Leu Tyr Xaa Xaa Pro Pro Ser Glu Cys Val His Ile  
       20                              25                              30

Ser Ser Gly Leu Pro  
       35

## (2) INFORMATION FOR SEQ ID NO: 255:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide

(B) LOCATION: -23..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4  
seq LHCLLVFILVEF/CK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

Met Ser Ala Phe Tyr Leu Ser Tyr Ser Leu Leu His Cys Leu Leu Ile  
-20 -15 -10  
Val Phe Ile Leu Val Glu Phe Cys Lys Lys Leu Thr Tyr Phe  
-5 1 5

(2) INFORMATION FOR SEQ ID NO: 256:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -30..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4  
seq SAGVVLTM DGASA/EQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

Met Ala Glu Ala Lys Leu Val Gln Gly Ser Leu Val Ala Pro Gln Arg  
-30 -25 -20 -15  
Xaa Ser Ala Gly Val Val Leu Thr Met Asp Gly Ala Ser Ala Glu Gln  
-10 -5 1  
Asp Gly Leu Gln Glu Asp Arg Ser His Ser Gly Pro Ser Ser Leu Pro  
5 10 15  
Glu Ala His Arg  
20

(2) INFORMATION FOR SEQ ID NO: 257:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: -59..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4  
seq VLLTISTNASVLG/DG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

Met Lys Gly Val Gly Pro Glu Gln Leu Asn Asp Gly Ala Pro Ser Asn  
                    -55                    -50                    -45

Glu Ile Glu Met Thr Pro Cys Phe Phe Ser Glu Phe Leu Leu Leu Asp  
                    -40                    -35                    -30

Val Gly Val Val Asn Ile Val Val Ile Lys Met Ser Tyr Asn Val Leu  
                    -25                    -20                    -15

Leu Thr Ile Ser Thr Asn Ala Ser Val Leu Gly Asp Gly Ala His Arg  
                    -10                    -5                    1                    5

Val Thr Thr Arg Ile Arg Arg Pro Gly Gly  
                    10                    15

## (2) INFORMATION FOR SEQ ID NO: 258:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: -45..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.8  
seq QLFWVTASTFCRS/DI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

Met Leu Arg Lys Leu Ser Ala Ser Asn Glu Asn Leu Cys Leu Leu Ser  
-45                    -40                    -35                    -30

Asn Pro Ser His Asn Glu Val Tyr Leu Ile Arg Cys Cys Glu Ser His  
                    -25                    -20                    -15

Gln Leu Phe Trp Val Thr Ala Ser Thr Phe Cys Arg Ser Asp Ile Ala  
-10 -5 1  
Thr Met Ala Ser Leu Leu Pro Ser Val Leu Leu Met Gln Leu Phe Ser  
5 10 15  
Thr Phe Phe Leu Asn Leu  
20 25

## (2) INFORMATION FOR SEQ ID NO: 259:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -15..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq PLILLPLNPFVLQ/VA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

Met Tyr Pro Leu Ile Leu Leu Pro Leu Asn Pro Phe Val Leu Gln Val  
-15 -10 -5 1  
Ala Gly

## (2) INFORMATION FOR SEQ ID NO: 260:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -23..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq RGFVAVGLGQISA/SP



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

```

Met Leu Leu Arg Pro Ser Pro Gly Ser Pro Arg Gly Phe Val Ala Val
      -20                      -15                      -10

Gly Leu Gly Gln Ile Ser Ala Ser Pro Ser Met Ala Cys Lys Leu Thr
      -5                      1                      5

Ile Leu Gln His Thr Ser Val Phe Arg Val Val Val Phe Arg Thr Pro
  10                      15                      20                      25

Leu Val Arg Gly Pro Leu Ser Arg Ser Asn Glu Leu Trp Leu His His
      30                      35                      40

Leu Ser Ser

```

(2) INFORMATION FOR SEQ ID NO: 261:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -18..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq ATACGPAAHQCSA/VP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

```

Met Ala Arg Pro Gly Ala Thr Ala Cys Gly Pro Ala Ala His Gln Cys
      -15                      -10                      -5

Ser Ala Val Pro Leu Trp Ser Pro Gly
      1                      5

```

(2) INFORMATION FOR SEQ ID NO: 262:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq LSLCIXXLEHLFT/WP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

Met Glu Pro Val Ser Ser Leu Ser Leu Cys Ile Xaa Xaa Leu Glu His  
                  -15                  -10                  -5

Leu Phe Thr Trp Pro Lys Gly  
                  1

(2) INFORMATION FOR SEQ ID NO: 263:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq WCSAAAWRSPLSA/AT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

Met Arg Pro Ala Gly Arg Trp Cys Ser Ala Ala Ala Trp Arg Ser Pro  
                  -15                  -10                  -5

Leu Ser Ala Ala Thr Leu Lys Cys Pro Leu Arg Gly  
                  1                                  5

(2) INFORMATION FOR SEQ ID NO: 264:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -16..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.7  
seq CAYVLFFFNGCLY/RR

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

Met Trp Leu Cys Ala Tyr Val Leu Phe Phe Phe Asn Gly Cys Leu Tyr  
-15 -10 -5

Arg Arg Lys  
1

## (2) INFORMATION FOR SEQ ID NO: 265:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -15..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.6  
seq LLHRAVVLRLQQA/CR

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

Met Leu Leu Leu His Arg Ala Val Val Leu Arg Leu Gln Gln Ala Cys  
-15 -10 -5 1

Arg Pro Thr Ser Leu Pro Asp Ser Ser Gln Ser Pro Gln Gly Ser Ala  
5 10 15

Phe Arg Pro Ala Pro Gln Met Ile His Phe Ser Pro Leu Xaa  
20 25 30

## (2) INFORMATION FOR SEQ ID NO: 266:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 53 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -32..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 3.6  
seq SLVPSMCFHVTNS/IK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

Met Glu Met Phe Gly Xaa Xaa Glu Lys Asp Phe Ser Ser Val Glu Gly  
-30 -25 -20  
Val Leu Xaa Ser Leu Val Pro Ser Met Cys Phe His Val Thr Asn Ser  
-15 -10 -5  
Ile Lys Met Pro Trp Phe Pro Ser Gln Pro Gly Thr Cys Thr Gln Lys  
1 5 10 15  
Asp Cys Pro Pro Lys  
20

(2) INFORMATION FOR SEQ ID NO: 267:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 amino acids

(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -46..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 3.6  
seq LLGVHASFQMSVA/AR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

Met Gln Met His Gly Trp Arg Trp Asp Pro His Ser Ser Glu Gln Leu  
-45 -40 -35  
Asp Leu Ala His Thr Leu Ser Arg Glu Ala Ser Leu Glu Asn Asn Thr  
-30 -25 -20 -15  
Ala Leu Leu Gly Val His Ala Ser Phe Gln Met Ser Val Ala Ala Arg

-10

-5

1

## (2) INFORMATION FOR SEQ ID NO: 268:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -34..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.6  
seq VGTGVLTSRLARA/TP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

Met Ala Ser Pro Arg Gly Thr Asp Tyr Asn Gln Thr Pro Asn Thr Thr  
                  -30                  -25                  -20

Met Tyr Cys Tyr Ala Val Gly Thr Gly Val Leu Thr Ser Arg Leu Ala  
                  -15                  -10                  -5

Arg Ala Thr Pro Gly  
                  1

## (2) INFORMATION FOR SEQ ID NO: 269:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -42..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.6  
seq LHCLCPFPALFLS/VT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

Met Ala Pro Ile Leu Ser Ser Phe Lys Ser Leu Leu Lys Tyr His Leu  
           -40                              -35                              -30

Leu Glu Thr Ser Leu Ser Ile Leu Leu Lys Pro Val Thr Leu His Cys  
       -25                              -20                              -15

Leu Cys Pro Phe Pro Ala Leu Phe Leu Ser Val Thr Phe Ile Tyr Leu  
       -10                              -5                              1                              5

Thr Tyr Tyr Ile Phe Asn Leu Tyr Ile Leu Phe Ile Val Cys Leu Leu  
                   10                              15                              20

Tyr Trp Asn Val Leu Ser Met  
                   25

## (2) INFORMATION FOR SEQ ID NO: 270:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -22..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq ILVPWWLPPFVYT/AI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

Met Asn Arg Leu Ser Lys His Leu Ile Ile Leu Val Pro Trp Trp Leu  
       -20                              -15                              -10

Pro Pro Phe Val Tyr Thr Ala Ile Ser Tyr Val Gln Leu Pro Gly  
       -5                              1                              5

## (2) INFORMATION FOR SEQ ID NO: 271:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -29..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq ALITILILYSSNS/AI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

```

Met Ser Ser Asn Lys Glu Gln Arg Ser Ala Val Phe Val Ile Leu Phe
      -25                -20                -15

Ala Leu Ile Thr Ile Leu Ile Leu Tyr Ser Ser Asn Ser Ala Ile Gly
      -10                -5                1

```

## (2) INFORMATION FOR SEQ ID NO: 272:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -68..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq CLFLSPQSFVLVS/WA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

```

Met Asp Met Lys Ser Asn Thr Gly His Gly Leu Phe Leu Gly Arg Gln
      -65                -60                -55

Pro Ser Phe Ser Val Arg Ser Met Pro Gly Thr Pro Ala Leu Ala Ile
      -50                -45                -40

Cys Gln Pro His Asn Pro Gly Pro Pro Met Gly Thr Pro Thr Glu Asp
      -35                -30                -25

Pro Ser Gly Cys Ser Phe Pro Cys Leu Phe Leu Ser Pro Gln Ser Phe
      -20                -15                -10                -5

Leu Val Leu Ser Trp Ala Ile Ser Arg
      1                5

```

## (2) INFORMATION FOR SEQ ID NO: 273:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -29..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq LSLSSTLLLTSHH/HQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

Met Ser Glu Ala Gly Cys Lys Pro Ser Arg Pro Glu His Gly Ser Phe  
                  -25                  -20                  -15

Leu Ser Leu Ser Ser Thr Leu Leu Leu Thr Ser His His His Gln Ser  
                  -10                  -5                  1

Ser Asp Phe Gly  
                  5

(2) INFORMATION FOR SEQ ID NO: 274:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 114 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 12.8  
seq XVFLVALLRGVQC/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

Met Glu Ser Gly Xaa Gly Xaa Val Phe Leu Val Ala Leu Leu Arg Gly  
                  -15                  -10                  -5

Val Gln Cys Gln Val Gln Ile Val Gln Ser Gly Gly Gly Val Val Gln  
                  1                  5                  10

Pro Gly Lys Ser Gln Thr Leu Ser Cys Val Thr Tyr Gly Phe Arg Phe



15				20				25							
Asp 30	Asp	Phe	Gly	Phe	His 35	Trp	Val	Arg	Gln	Ala 40	Pro	Gly	Lys	Gly	Leu 45
Glu	Trp	Val	Ala	Met 50	Ile	Arg	Tyr	Asp	Gly 55	Ser	Asn	Lys	Phe	Tyr 60	Ser
Lys	Ser	Val	Gln 65	Gly	Arg	Phe	Leu	Ile 70	Ser	Arg	Asp	Asn	Ser 75	Arg	Asn
Gln	Val	Tyr 80	Leu	Ser	Leu	Asn	Arg 85	Leu	Arg	Val	Asp	Asp 90	Thr	Ala	Val
Tyr	Tyr 95														

(2) INFORMATION FOR SEQ ID NO: 275:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: -17..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 9.3  
seq LFTLLLLQSLLLG/CC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

Met Leu Cys Arg Leu Phe Thr Leu Leu Leu Leu Gln Ser Leu Leu Leu  
-15 -10 -5

Gly Cys Cys Ile Tyr Xaa Pro Gly Asn Gly  
1 5

(2) INFORMATION FOR SEQ ID NO: 276:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 97 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -26..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.3  
seq FLLLVAGPRWVLS/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

Met Asp Leu Leu His Lys Asn Met Lys His Leu Trp Phe Phe Leu Leu  
-25 -20 -15

Leu Val Ala Gly Pro Arg Trp Val Leu Ser Gln Val Arg Leu Glu Gln  
-10 -5 1 5

Trp Gly Ser Gly Leu Val Lys Ser Ser Glu Thr Leu Ser Leu Thr Cys  
10 15 20

Ala Val Tyr Gly Gly Ser Ala Ile Ser Asp Tyr Trp Ala Trp Ile Arg  
25 30 35

Gln Phe Pro Gly Lys Gly Val Glu Trp Ile Gly Glu Ile Asn His Ser  
40 45 50

Gly Ala Thr His Tyr Ile Arg Pro Ser Gly Val Glu Ser Pro Ser Pro  
55 60 65 70

Leu

(2) INFORMATION FOR SEQ ID NO: 277:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 77 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -51..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.1  
seq VCLCGTFCFPCLG/CQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

Met Gln Ala Gln Ala Pro Val Val Val Val Thr Gln Pro Gly Val Gly  
-50 -45 -40

Pro Gly Pro Ala Pro Gln Asn Ser Asn Trp Gln Thr Gly Met Cys Asp

```

-35          -30          -25          -20
Cys Phe Ser Asp Cys Gly Val Cys Leu Cys Gly Thr Phe Cys Phe Pro
      -15          -10          -5
Cys Leu Gly Cys Gln Val Ala Ala Asp Met Asn Glu Cys Cys Leu Cys
      1          5          10
Gly Thr Ser Val Ala Met Arg Thr Leu Xaa Arg Xaa Arg
      15          20          25

```

## (2) INFORMATION FOR SEQ ID NO: 278:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -18..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.7  
seq LLLLPVLGLLVSS/KT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

```

Met Lys Ala Leu Cys Leu Leu Leu Leu Pro Val Leu Gly Leu Leu Val
      -15          -10          -5
Ser Ser Lys Thr Leu Cys Ser Met Glu Glu Ala Ile Asn Glu Arg Ile
      1          5          10
Gln Glu Val Ala Gly Ser Leu Ile Phe Arg Ala Ile Ser Ser Ile Gly
      15          20          25          30
Arg Gly Ser Glu

```

## (2) INFORMATION FOR SEQ ID NO: 279:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -21..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.6  
seq LLXIVGLXLPTXG/QX

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

```

Met Ser Pro Ser Gly Arg Leu Cys Leu Leu Xaa Ile Val Gly Leu Xaa
-20                               -15                -10

Leu Pro Thr Xaa Gly Gln Xaa Leu Lys Asp Thr Xaa Ser Ser Ser Ser
-5                               1                    5                10

Ala Asp Ser Thr Ile Met Asp Ile Gln Val Pro Thr Arg Ala Pro Asp
15                               20                25

Ala Val Tyr Thr Glu Leu Gln Pro Thr His Gly
30                               35

```

## (2) INFORMATION FOR SEQ ID NO: 280:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -21..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.2  
seq FLVSNMLLA EAYG/SG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

```

Met Leu Leu Ala Trp Val Gln Ala Phe Leu Val Ser Asn Met Leu Leu
-20                               -15                -10

Ala Glu Ala Tyr Gly Ser Gly Gly Cys Phe Trp Asp Asn Gly His Leu
-5                               1                    5                10

Tyr Arg Glu Asp Gln Thr Ser Pro Ala Pro Gly Leu Arg Cys Leu Asn
15                               20                25

Trp Leu Asp Ala Gln Asn Gly Leu
30                               35

```

## (2) INFORMATION FOR SEQ ID NO: 281:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -62...-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.8  
seq LXLTCVSGGSIS/RT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

```

Met Leu Ser Glu Ser Arg Gly Pro Pro Val Gln Glu His Glu Ala Pro
  -60                      -55                      -50

Val Val Leu Pro Pro Ala Gly Gly Gly Ser Gln Met Gly Pro Val Pro
  -45                      -40                      -35

Ala Ala Xaa Ala Gly Glu Ser Gly Pro Gly Xaa Val Lys Pro Leu Glu
-30                      -25                      -20                      -15

Thr Leu Xaa Leu Thr Cys Ser Val Ser Gly Gly Ser Ile Ser Arg Thr
      -10                      -5                      1

Ser Phe Tyr Trp Gly Trp Ile Arg Gln Pro Pro Gly Lys Gly Leu Glu
      5                      10                      15

Trp Ile Gly Ser Ile Tyr Asp Xaa Gly Ser Thr Tyr Tyr Asn Pro Ser
  20                      25                      30

Leu Xaa Xaa Xaa Val Thr Ile Ser Val Asp Thr Ser Lys Asn Gln Val
  35                      40                      45                      50

Ser Leu Lys Val Ser Ser Val Thr Ala Ala Asp Thr Ala Val Tyr His
      55                      60                      65

Cys Ala Arg Gly
      70

```

## (2) INFORMATION FOR SEQ ID NO: 282:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 137 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: -108..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.3  
seq ACMTLTASPGVFP/SL

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

```

Met Thr Ser Gly Gln Ala Arg Ala Ser Xaa Gln Ser Pro Gln Ala Leu
      -105                -100                -95

Glu Asp Ser Gly Pro Val Asn Ile Ser Val Ser Ile Thr Leu Thr Leu
      -90                -85                -80

Asp Pro Leu Lys Pro Phe Gly Gly Tyr Ser Arg Asn Val Thr His Leu
      -75                -70                -65

Tyr Ser Thr Ile Leu Gly His Gln Ile Gly Leu Ser Gly Arg Glu Ala
      -60                -55                -50                -45

His Glu Glu Ile Asn Ile Thr Phe Thr Leu Pro Thr Ala Trp Ser Ser
      -40                -35                -30

Asp Asp Cys Ala Leu His Gly His Cys Glu Gln Val Val Phe Thr Ala
      -25                -20                -15

Cys Met Thr Leu Thr Ala Ser Pro Gly Val Phe Pro Ser Leu Tyr Ser
      -10                -5                1

His Arg Thr Val Phe Leu Thr Arg Thr Ala Thr Pro Arg Ser Gly Thr
      5                10                15                20

Arg Ser Ser Gln Leu Pro Glu Met Pro
      25

```

## (2) INFORMATION FOR SEQ ID NO: 283:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: -21..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 5.1  
seq LLLKIWLLQRPES/QE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

```

Met Leu Gly Gly Asp His Arg Ala Leu Leu Leu Lys Ile Trp Leu Leu
-20                               -15                               -10

Gln Arg Pro Glu Ser Gln Glu Gly Leu Leu Pro Gly Arg Leu Val Val
-5                               1                               5                               10

Met Glu Arg Arg Val Lys Asn Asp Leu Met Ser Phe Leu Ser Thr Ala
15                               20                               25

```

(2) INFORMATION FOR SEQ ID NO: 284:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -29..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.6  
seq SLMSLLDESSCQA/VG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

```

Met Arg Phe Arg Lys Ala Trp Ala Pro Val Leu Ala Ala Leu Ser His
-25                               -20                               -15

Ser Leu Met Ser Leu Leu Asp Glu Ser Ser Cys Gln Ala Val Gly Arg
-10                               -5                               1

Pro Val Glu Lys Leu Ala Arg Asn Trp Trp Gly Pro Phe Pro Pro Ile
5                               10                               15

Ala Ser Lys Glu Leu Asn Pro Ala
20                               25

```

(2) INFORMATION FOR SEQ ID NO: 285:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 105 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -82..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 4.4  
seq NLPHLQVVGLTWG/HI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

```
Met Tyr Val Trp Pro Cys Ala Val Val Leu Ala Gln Tyr Leu Trp Phe
    -80                      -75                      -70

His Arg Arg Ser Leu Pro Gly Lys Ala Ile Leu Glu Ile Gly Ala Gly
    -65                      -60                      -55

Val Ser Leu Pro Gly Ile Leu Ala Ala Lys Cys Gly Ala Glu Val Ile
    -50                      -45                      -40                      -35

Leu Ser Asp Ser Ser Glu Leu Pro His Cys Leu Glu Val Cys Arg Gln
                -30                      -25                      -20

Ser Cys Gln Met Asn Asn Leu Pro His Leu Gln Val Val Gly Leu Thr
                -15                      -10                      -5

Trp Gly His Ile Ser Trp Asp Leu Leu Ala Leu Pro Pro Gln Asp Ile
    1                      5                      10

Ile Leu Ala Ser Asp Val Phe Phe Glu
    15                      20
```

(2) INFORMATION FOR SEQ ID NO: 286:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 126 amino acids

(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -56..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 4.4  
seq LWKLALQSSSCLS/LF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:



```

Met Leu Asn Pro Ala Gln Xaa Asp Thr Met Pro Cys Glu Tyr Leu Ser
-55                               -50                               -45

Leu Asp Ala Met Glu Lys Trp Ile Ile Phe Gly Phe Ile Leu Cys His
-40                               -35                               -30                               -25

Gly Ile Leu Asn Thr Xaa Ala Thr Ala Leu Asn Leu Trp Lys Leu Ala
-20                               -15                               -10

Leu Gln Ser Ser Ser Cys Leu Ser Leu Phe Arg Asp Glu Val Phe His
-5                               1                               5

Ile His Lys Ala Ala Glu Asp Leu Phe Val Asn Ile Arg Gly Tyr Asn
10                               15                               20

Lys Arg Ile Asn Asp Ile Arg Glu Cys Lys Xaa Ala Ala Val Ser His
25                               30                               35                               40

Ala Gly Ser Met His Arg Glu Arg Arg Lys Xaa Leu Arg Ser Ala Leu
45                               50                               55

Lys Glu Leu Ala Thr Val Leu Ser Asp Gln Pro Gly Leu Leu
60                               65                               70

```

## (2) INFORMATION FOR SEQ ID NO: 287:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -24..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.9  
seq GVCLSVPSLPSIS/RP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

```

Met Asn Ala Gln Ala Ser Ser Ser Arg Cys His Gly Val Cys Leu Ser
-20                               -15                               -10

Val Pro Ser Leu Pro Ser Ile Ser Arg Pro Pro
-5                               1

```

## (2) INFORMATION FOR SEQ ID NO: 288:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 74 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -67..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq QVLDSVLVGPVPA/ER

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

```
Met Ala Lys Val Gln Val Asn Asn Val Val Val Leu Asp Asn Pro Ser
    -65                      -60                      -55

Pro Phe Tyr Asn Pro Phe Gln Phe Glu Ile Thr Phe Glu Cys Ile Glu
    -50                      -45                      -40

Asp Leu Ser Glu Asp Leu Glu Trp Lys Ile Ile Tyr Val Gly Ser Ala
    -35                      -30                      -25                      -20

Glu Ser Glu Glu Tyr Asp Gln Val Leu Asp Ser Val Leu Val Gly Pro
    -15                      -10                      -5

Val Pro Ala Glu Arg His Met Phe Val Phe
    1                      5
```

(2) INFORMATION FOR SEQ ID NO: 289:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -21..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.7  
seq ETCALASHSGSSG/SK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

```
Met Ala Asp Val Glu Asp Gly Glu Glu Thr Cys Ala Leu Ala Ser His
```

-20

-15

-10

Ser Gly Ser Ser Gly Ser Lys Ser Gly Gly Asp Lys Met Phe Ser Leu  
-5 1 5 10

Lys Lys Trp Asn Ala Val  
15

## (2) INFORMATION FOR SEQ ID NO: 290:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -14..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq WTCLLGDCGPPEA/FT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

Met Trp Thr Cys Leu Leu Gly Asp Cys Gly Pro Pro Glu Ala Phe Thr  
-10 -5 1

Ser Tyr Gln Pro Pro Arg  
5

## (2) INFORMATION FOR SEQ ID NO: 291:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 11.7  
seq VFCLLAVAPGAHS/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

```

Met Asp Trp Thr Trp Xaa Val Phe Cys Leu Leu Ala Val Ala Pro Gly
           -15                -10                -5

Ala His Ser Gln Val Gln Leu Val Gln Ser Xaa Ala Xaa Val Arg Xaa
           1                   5                   10

Pro Gly Ala Ser Val Lys Val Ser Cys Lys Pro Ser Gly Tyr Ser Phe
   15                20                25

Thr Ser His Tyr Val His Trp Val Arg Xaa Asp Pro Gly Gln Xaa Leu
  30                35                40                45

Glu Trp Met Gly Asp Gly
           50

```

(2) INFORMATION FOR SEQ ID NO: 292:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 75 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -33..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.9  
seq LVSLLLLLLTRVQP/GT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

```

Met Asp Asn Ser Trp Arg Leu Gly Pro Ala Ile Gly Leu Ser Ala Gly
           -30                -25                -20

Gln Ser Gln Leu Leu Val Ser Leu Leu Leu Leu Thr Arg Val Gln
   -15                -10                -5

Pro Gly Thr Asp Val Ala Ala Pro Glu His Ile Ser Tyr Val Pro Gln
   1                   5                   10                15

Leu Ser Asn Asp Thr Leu Ala Gly Arg Leu Thr Leu Ser Thr Phe Thr
   20                25                30

Leu Glu Gln Pro Leu Gly Gln Phe Ser Ser Arg
   35                40

```

(2) INFORMATION FOR SEQ ID NO: 293:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 95 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganqlia

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig peptide

(B) LOCATION: -19...-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 10.9  
seq FLLLVAAPRWVLS/OV

seq FLLVAAPRWVLS/OV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

Met	Xaa	His	Leu	Xaa	Phe	Phe	Leu	Leu	Leu	Val	Ala	Ala	Pro	Arg	Trp
				-15					-10					-5	
Val	Leu	Ser	Gln	Val	Leu	Leu	Gln	Glu	Ser	Gly	Pro	Glu	Leu	Val	Lys
			1				5					10			
Pro	Ser	Xaa	Thr	Leu	Ser	Leu	Thr	Xaa	Ala	Val	Ser	Gly	Gly	Ser	Ile
	15					20					25				
Ser	Gly	Gly	Pro	Tyr	Tyr	Trp	Asn	Trp	Val	Xaa	Gln	His	Pro	Gly	Lys
30					35					40					45
Gly	Leu	Glu	Xaa	Ile	Gly	Asn	Ile	Tyr	Tyr	Asn	Gly	Ser	Thr	Phe	Xaa
				50					55					60	
Xaa	Pro	Val	Pro	Gln	Xaa	Ser	Xaa	Tyr	His	Ile	Xaa	Arg	Arg	Arg	
			65					70					75		

(2) INFORMATION FOR SEQ ID NO: 294:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 85 amino acids

(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig peptide

(B) LOCATION: -28...-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 9.6  
seq LLTLLLGLTEVAG/EE

seq LLTLLGLTEVAG/EE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

```

Met Pro Val Pro Ala Ser Trp Pro His Pro Pro Gly Pro Phe Leu Leu
      -25                -20                -15

Leu Thr Leu Leu Leu Gly Leu Thr Glu Val Ala Gly Glu Glu Leu
      -10                -5                1

Gln Met Ile Gln Pro Glu Lys Leu Leu Leu Val Thr Val Gly Lys Thr
  5                10                15                20

Ala Thr Leu His Cys Thr Val Thr Ser Leu Leu Pro Val Gly Pro Val
                25                30                35

Leu Trp Phe Arg Gly Val Gly Pro Gly Arg Glu Leu Ile Tyr Asn Gln
                40                45                50

Lys Glu Gly Leu Xaa
      55

```

## (2) INFORMATION FOR SEQ ID NO: 295:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -15..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.6  
seq EYVLLFLALCSA/KP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

```

Met Lys Glu Tyr Val Leu Leu Leu Phe Leu Ala Leu Cys Ser Ala Lys
-15                -10                -5                1

Pro Phe Phe Ser Pro Ser His Ile Ala Leu Lys Asn Met Met Leu Lys
      5                10                15

Asp Met Glu Asp Thr Glu
      20

```

## (2) INFORMATION FOR SEQ ID NO: 296:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -17..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.3  
seq LALSLILVLAFG/IP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

Met Ala Gln Ser Leu Ala Leu Ser Leu Leu Ile Leu Val Leu Ala Phe  
-15 -10 -5  
Gly Ile Pro Arg Thr Gln Gly Ser Asp Gly Gly Ala Gln Asp Cys Cys  
1 5 10 15  
Leu Lys Tyr Ser Gln Thr Arg  
20

(2) INFORMATION FOR SEQ ID NO: 297:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 83 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -17..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.2  
seq LLLITAILAVAVG/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

Met Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val  
-15 -10 -5  
Gly Phe Pro Val Ser Gln Asp Xaa Glu Arg Glu Lys Arg Ser Ile Ser  
1 5 10 15  
Asp Ser Asp Glu Leu Ala Ser Gly Phe Phe Val Phe Pro Tyr Pro Tyr

20 25 30

Pro Phe Arg Pro Leu Pro Pro Ile Pro Phe Pro Arg Phe Pro Trp Phe  
35 40 45

Arg Arg Asn Phe Pro Ile Pro Ile Pro Glu Ser Ala Pro Thr Thr Pro  
50 55 60

Leu Pro Met  
65

## (2) INFORMATION FOR SEQ ID NO: 298:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -17..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.2  
seq LLLITAILAVAVG/FP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

Met Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val  
-15 -10 -5

Gly Phe Pro Val Ser Gln Asp Gln Glu Arg Glu Lys Arg Ser Ile Ser  
1 5 10 15

Asp Ser Asp Glu Leu Ala Ser Gly Xaa Phe Val Phe Pro Tyr Pro Tyr  
20 25 30

Pro Phe Arg Pro Leu Pro Pro Ile Pro Phe Pro Arg Phe Pro Trp Phe  
35 40 45

Arg Arg Xaa Phe Pro Ile Pro Ile Pro Glu Ser Ala Pro Gly  
50 55 60

## (2) INFORMATION FOR SEQ ID NO: 299:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN



## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Placenta

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: -18..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 8.1  
seq LFTAILAFSLAQS/FG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

Met Arg Ile Met Leu Leu Phe Thr Ala Ile Leu Ala Phe Ser Leu Ala  
-15 -10 -5  
Gln Ser Phe Gly Ala Val Cys Lys Glu Pro Gln Glu Glu Val Val Pro  
1 5 10  
Gly Gly Gly Arg Ser Lys Arg Asp Pro Asp Leu Tyr Gln Leu Leu Gln  
15 20 25 30  
Arg Pro Trp

## (2) INFORMATION FOR SEQ ID NO: 300:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 125 amino acids  
(B) TYPE: AMINO ACID  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: -17..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 7.9  
seq VLLLGLLSHCTVS/VS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

Met Ala Trp Thr Val Leu Leu Leu Gly Leu Leu Ser His Cys Thr Val  
-15 -10 -5  
Ser Val Ser Ser Tyr Glu Leu Thr Gln Pro Pro Ser Val Ser Val Ala  
1 5 10 15  
Pro Gly Glu Thr Ala Thr Ile Ser Cys Gly Ala Asn Asn Val Gly Arg  
20 25 30  
Lys Asn Val Gln Trp Tyr Gln Gln Lys Ala Gly Gln Ala Pro Val Leu

35                      40                      45  
 Val Ile Tyr His Asp Val Glu Arg Pro Ser Gly Ile Pro Glu Arg Phe  
           50                      55                      60  
 Ser Gly Ser Asn Ser Gly Ser Pro Ala Lys Leu Thr Ile Ser Arg Val  
           65                      70                      75  
 Glu Ala Gly Asp Glu Ala Asp Tyr Xaa Cys Xaa Val Trp Asp Ser Asp  
           80                      85                      90                      95  
 Ser Asp His Thr Val Ile Phe Gly Gly Gly Thr Lys Leu  
                           100                      105

## (2) INFORMATION FOR SEQ ID NO: 301:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 82 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -58..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.7  
seq PXLLLQTLPASTX/XP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

Met Thr Ile Leu His Thr Gly Xaa Asn Pro Phe Arg Pro Ser Gln Arg  
           -55                      -50                      -45  
 Trp Thr Ala Pro Ala Leu Leu His His Arg Pro Xaa Thr Xaa Pro Pro  
           -40                      -35                      -30  
 Ser Xaa His Arg Ser Arg Cys Thr Glu Xaa Val Gly Ile Pro Xaa Leu  
           -25                      -20                      -15  
 Leu Leu Gln Thr Leu Pro Ala Ser Thr Xaa Xaa Pro Gln Ala Phe Arg  
           -10                      -5                      1                      5  
 Arg Xaa Ser Asp Pro Pro Ala Lys Pro Pro Gln Ile Tyr Tyr Arg Val  
           10                      15                      20  
 Gln His

## (2) INFORMATION FOR SEQ ID NO: 302:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 115 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -20..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.2  
seq LLLVAAAPKXXLS/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

```
Met Lys His Leu Trp Phe Phe Leu Leu Leu Leu Val Ala Ala Pro Lys
-20          -15          -10          -5

Xaa Xaa Leu Ser Gln Val Gln Leu Arg Glu Ser Gly Pro Gly Leu Val
      1              5              10

Glu Pro Ser Gln Thr Leu Ser Leu Thr Cys Ser Val Ser Arg Gly Ser
      15              20              25

Val Asn Ser Gly Gly Tyr Tyr Trp Ser Trp Ile Arg Gln His Pro Gly
      30              35              40

Lys Gly Leu Glu Trp Ile Gly Tyr Val Tyr Tyr Gly Gly Xaa Thr Tyr
      45              50              55              60

Tyr Asn Pro Ser Leu Lys Ser Arg Val Thr Leu Ser Ala Asp Thr Ser
      65              70              75

Lys Asn Gln Phe Phe Leu Arg Leu Thr Ser Met Thr Ala Ala Asp Thr
      80              85              90

Ala Ser Gly
      95
```

(2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide

(B) LOCATION: -21..-1  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 6.9  
 seq LVCGSLGLSNVSG/IY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

```
Met Leu Ser Tyr Phe Leu Ser Ser Leu Val Cys Gly Ser Leu Gly Leu
-20                               -15                -10

Ser Asn Val Ser Gly Ile Tyr Asp Ser Lys Lys Lys Arg Lys Thr Gly
-5                               1                5                10

Ala Phe Arg Thr Gln Leu Phe Trp Gly Val Gly
15                               20
```

(2) INFORMATION FOR SEQ ID NO: 304:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 57 amino acids  
 (B) TYPE: AMINO ACID  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: -42..-1  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 6.4  
 seq ELPALALLHAGHA/EP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

```
Met Gly Thr Gln Asp Pro Gln Ala Glu Gln Gly Leu Arg Ile Pro Leu
-40                               -35                -30

Pro Gly Leu Leu Leu Ser Lys His His His Pro Ala Pro Glu Leu Pro
-25                               -20                -15

Ala Leu Ala Leu Leu His Ala Gly His Ala Glu Pro Ala Gln Asp Gly
-10                               -5                1                5

Glu Pro Gly His Pro Arg Gly Pro Gly
10                               15
```

(2) INFORMATION FOR SEQ ID NO: 305:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 76 amino acids  
 (B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -33..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 6.3  
seq SXXPLXSVQLXHA/QR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

Met Met Thr Ile Tyr Ala Leu Ser Asn Glu Phe Ala Phe Lys Ile Asn  
-30 -25 -20

Glu Glu Gln Leu Ser Xaa Xaa Pro Leu Xaa Ser Val Gln Leu Xaa His  
-15 -10 -5

Ala Gln Arg Phe Leu Leu Asp Ser Ser Trp Ser Gly Val Ile Pro Phe  
1 5 10 15

Phe Phe Ser Cys Ser Cys Leu Pro Phe Leu Tyr Pro Pro Lys Trp Arg  
20 25 30

Gln Ile His Asp Leu Lys Asp Thr Gln Tyr Arg Ser  
35 40

(2) INFORMATION FOR SEQ ID NO: 306:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Lymphocytes

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: -21..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 6.1  
seq LEMLTAFASHIRA/RD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

Met Arg Gly Ala His Leu Xaa Ala Leu Glu Met Leu Thr Ala Phe Ala  
-20 -15 -10

Ser His Ile Arg Ala Arg Asp Ala Ala Arg  
 -5 1 5

## (2) INFORMATION FOR SEQ ID NO: 307:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -52..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.1  
seq IILLIHTMQVCTT/HP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

Met Asn Pro Glu Ser Pro Gln Gln Leu Glu Arg Gln Ser Thr Gly Pro  
 -50 -45 -40

Arg Thr Gly Thr Arg Arg Cys Leu Ser Lys Phe Thr Trp Cys Thr Ser  
 -35 -30 -25

Arg Met Met Thr Gln Thr Cys Ile Ile Leu Leu Ile His Thr Met Gln  
 -20 -15 -10 -5

Val Cys Thr Thr His Pro Thr Val Leu Ser His Thr Leu Leu Gln Arg  
 1 5 10

Pro Lys Pro Thr Asp Pro Arg  
 15

## (2) INFORMATION FOR SEQ ID NO: 308:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1

(C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 6.1  
 seq IILLIHTMQVCTT/HP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

Met Met Thr Gln Thr Cys Ile Ile Leu Leu Ile His Thr Met Gln Val  
                   -15                  -10                  -5  
 Cys Thr Thr His Pro Thr Val Leu Ser His Thr Leu Leu Gln Arg Pro  
                   1                  5                  10  
 Met

(2) INFORMATION FOR SEQ ID NO: 309:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 82 amino acids  
 (B) TYPE: AMINO ACID  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: -25..-1  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 6  
 seq LLGLLVAVATVHL/VI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

Met Ala Gly Lys Gly Ser Ser Gly Arg Arg Pro Leu Leu Leu Gly Leu  
 -25                  -20                  -15                  -10  
 Leu Val Ala Val Ala Thr Val His Leu Val Ile Cys Pro Tyr Thr Lys  
                   -5                  1                  5  
 Val Glu Glu Ser Phe Asn Leu Gln Ala Thr His Asp Leu Leu Tyr His  
                   10                  15                  20  
 Trp Gln Asp Leu Glu Gln Tyr Asp His Leu Glu Phe Pro Gly Val Val  
                   25                  30                  35  
 Pro Arg Thr Xaa Leu Gly Pro Val Val Ile Ala Val Phe Ser Ser Pro  
                   40                  45                  50                  55  
 Ala Val

(2) INFORMATION FOR SEQ ID NO: 310:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 124 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -22..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9  
seq LIYILWQLTGSA/SG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

Met Ala Gly Ser Pro Thr Cys Leu Thr Leu Ile Tyr Ile Leu Trp Gln  
-20 -15 -10

Leu Thr Gly Ser Ala Ala Ser Gly Pro Val Lys Glu Leu Val Gly Ser  
-5 1 5 10

Val Gly Gly Ala Val Thr Phe Pro Leu Lys Ser Lys Val Lys Gln Val  
15 20 25

Asp Ser Ile Val Trp Thr Phe Asn Thr Thr Pro Leu Val Thr Ile Gln  
30 35 40

Pro Glu Gly Gly Xaa Ile Ile Val Thr Gln Asn Arg Asn Arg Glu Arg  
45 50 55

Val Asp Phe Pro Asp Gly Gly Tyr Ser Leu Lys Leu Ser Lys Leu Lys  
60 65 70

Lys Asn Asp Ser Xaa Ile Tyr Tyr Val Gly Ile Tyr Ser Ser Ser Leu  
75 80 85 90

Gln Gln Pro Xaa Thr Gln Glu Tyr Val Leu His Val  
95 100

## (2) INFORMATION FOR SEQ ID NO: 311:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: PROTEIN

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:



(A) NAME/KEY: sig\_peptide  
(B) LOCATION: -18..-1  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.8  
seq CFIILGLIICIOC/ST

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

```

Met Val Gly Met Val Cys Phe Ile Ile Leu Gly Leu Ile Ile Cys Ile
      -15                -10                -5

Gln Cys Ser Thr Gly
      1

```

(2) INFORMATION FOR SEQ ID NO: 312:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 116 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

```
(ix) FEATURE:
      (A) NAME/KEY: sig_peptide
      (B) LOCATION: -13..-1
      (C) IDENTIFICATION METHOD: Von Heijne matrix
      (D) OTHER INFORMATION:  score 5.7
                               seq MXLLHSLSSGVRA/PS
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

Met	Xaa	Leu	Leu	His	Ser	Leu	Ser	Ser	Gly	Val	Arg	Ala	Pro	Ser	Pro
	-		-10					-5					1		
Ala	Pro	Ser	Ser	Val	Pro	Leu	Gly	Ser	Glu	Lys	Pro	Ser	Asn	Val	Ser
	5					10					15				
Gln	Asp	Arg	Lys	Val	Pro	Val	Pro	Ile	Gly	Thr	Glu	Arg	Ser	Ala	Arg
20					25					30					35
Ile	Arg	Gln	Thr	Gly	Thr	Ser	Ala	Pro	Ser	Val	Ile	Gly	Ser	Asn	Leu
				40					45					50	
Ser	Thr	Ser	Val	Gly	His	Ser	Gly	Ile	Trp	Ser	Phe	Glu	Gly	Ile	Gly
			55					60					65		
Gly	Asn	Gln	Asp	Lys	Val	Asp	Trp	Cys	Asn	Pro	Gly	Met	Gly	Asn	Xaa
		70					75					80			
Met	Ile	His	Arg	Pro	Met	Ser	Asp	Pro	Gly	Val	Phe	Ser	Gln	His	Gln
	85					90					95				

Ala Thr Xaa Ala  
100

(2) INFORMATION FOR SEQ ID NO: 313:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -19..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.6  
seq PTLCVSSSPALWA/AS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

Met Thr Met Ala Glu Cys Pro Thr Leu Cys Val Ser Ser Ser Pro Ala  
                  -15                  -10                  -5

Leu Trp Ala Ala Ser Glu Thr Gly  
                  1                  5

(2) INFORMATION FOR SEQ ID NO: 314:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 133 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: -25..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.5  
seq AQLFACLLRLGTQ/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

Met Val Pro Leu Val Ala Val Val Ser Gly Pro Arg Ala Gln Leu Phe  
-25                  -20                  -15                  -10